Vol. 39, No. 2 April, 1949

Journal of Experimental Psychology

EDITED BY

FRANCIS W. IRWIN

UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CLARENCE H. GRAHAM, *Columbia University*
HARRY HELSON, *Bryn Mawr College*
DONALD B. LINDSLEY, *Northwestern University*
DONALD G. MARQUIS, *University of Michigan*
GEORGE A. MILLER, *Harvard University*
L. L. THURSTONE, *University of Chicago*
G. R. WENDT, *University of Rochester*
JOSEPH ZUBIN, *N. Y. Psychiatric Institute, Columbia University*

CONTENTS

Differential Sensitivity in Olfaction: BERNICE M. WENZEL 129
Concept Formation and Emergence of Contradictory Relations: J. C. DIXON 144
Work and Rest as Variables in Cyclical Motor Learning: G. A. KIMBLE AND E. A. BILODEAU ... 150
The Guessing-Sequence Hypothesis, The 'Spread of Effect,' and Number-Guessing Habits: W. O. JENKINS AND LETA M. CUNNINGHAM 158
Distributed Practice in Motor Learning: Progressively Increasing and Decreasing Rests: BARBARA S. COOK AND E. R. HILGARD 169
Control of Rest Interval Activities in Experiments on Reminiscence in Serial Verbal Learning: S. WITHEY, C. E. BUXTON, AND A. ELKIN 173
An Experimental Analysis of Latent Learning: J. P. SEWARD 177
Continuity in Human Learning: W. C. H. PRENTICE 187
Sensitization of the Beta-Response as a Function of the Wavelength of the Stimulus: D. A. GRANT, J. P. HORNSETH, AND H. W. HAKE 195
The Effect of Instructions upon Sensory Pre-Conditioning of Human Subjects: R. CHERNIKOFF AND W. J. BROGDEN 200
A Factorial Approach to the Analysis of Variances in Esthetic Judgments: J. P. GUILFORD AND J. W. HOLLEY 208
The Effects of Repeated Rotary Acceleration on the Oculo-Gyral Illusion: B. CLARK AND K. MACCORQUODALE 219
The Emergence of a Tonal Sensation: J. D. HARRIS AND C. K. MYERS 228
The Problem of Stimulation Deafness. III. The Functional and Histological Effects of a High-Frequency Stimulus: K. R. SMITH AND E. G. WEVER 238
The Projective Expression of Needs. IV. The Effect of the Need for Achievement on Thematic Apperception: D. C. MCCLELLAND, R. A. CLARK, T. B. ROBY, AND J. W. ATKINSON 242
Relative Self-Estimate and the Level of Aspiration: ALBERTA S. GILINSKY 256
The Effects of Punishment on Acts Motivated by Fear: G. T. GWINN 260
Accuracy of Visual Interpolation between Circular Scale Markers as a Function of the Separation between Markers: M. LEVZOREK 270

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. & R., authorized September 11, 1947.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published bi-monthly by the American Psychological Association, Inc. The annual subscription rate is $7.00 (foreign, $7.25), single copies, $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Articles of a length exceeding 25 printed pages cannot be accepted.

With rare exceptions, articles are published in the order of their receipt. Authors are supplied with 50 free offprints, without covers, in the case of articles of seven or more printed pages, and with five free copies of the issue in the case of shorter contributions. One-half of the cost of tables, cuts, formulas, and author's alterations in galley proof is charged to the author.

Immediate publication can be given to articles whose authors assume the full cost of publication.

Address articles submitted for publication to the editor, Francis W. Irwin, Department of Psychology, University of Pennsylvania, Philadelphia 4, Pa.

This JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. ANDERSON and W. L. VALENTINE, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, **41**, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts should be typed double-spaced throughout, *including tables, footnotes, and bibliography*. Tables, figures, and footnotes should be placed on separate pages at the end of the manuscript.

Journal of
Experimental Psychology

VOL. 39, NO. 2 APRIL, 1949

DIFFERENTIAL SENSITIVITY IN OLFACTION[1]

BY BERNICE M. WENZEL

Barnard College, Columbia University

INTRODUCTION AND BACKGROUND

The general purpose of the investigation reported in this paper is the measurement of differential sensitivity for an olfactory stimulus. Before such measurement could be made, a more specific problem had to be solved, viz., the design and construction of satisfactory apparatus for presenting the stimulus. This is a uniquely difficult problem in the field of olfaction, due to relative ignorance of the nature of the adequate stimulus. A summary of the previous attempts at solution and of the difficulties involved has been presented elsewhere by the present author (22). In this review it was shown that most of the previous techniques had utilized indirect units of stimulus intensity measurement. It was suggested that the most meaningful unit with our present state of knowledge would be the number of odorous molecules in the stimulus. The blast injection technique which eliminates voluntary inhalation and, hence, unmeasurable fluctuations in stimulus volume was considered a valuable addition to the field of olfactory investigation.

Only four reports have been published on aspects of relative discimination in olfaction. The most extensive, as well as the earliest, is that of Gamble (7) published in 1898. Her object was to investigate the applicability of Weber's Law to smell. For this purpose she selected 17 solids and 13 solutions as stimulus substances. The apparatus used was the Zwaardemaker olfactometer with some modifications for liquids. In order to obtain the DL's the method of just noticable differences was used. The most striking feature of Gamble's results is the pronounced variation among *S*s and among materials. Inspection of the limens reveals the small degree of consistency from substance to substance. Relative limens were obtained varying from less than $\frac{1}{6}$ to more than $\frac{1}{2}$ with a mode at $\frac{1}{3}$. Gamble concluded that Weber's Law is applicable to smell and that the limen lies between $\frac{1}{3}$ and $\frac{1}{4}$.

The experiment reported briefly by Toulouse and Vaschide in 1899 (19) had

[1] This article is substantially the same as the author's dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University. The writer is especially indebted to Prof. A. T. Poffenberger for constant assistance and encouragement as sponsor of the research. The writer wishes, also, to acknowledge special technical aid from H. E. Boorse, Professor of Physics in Barnard College, and from J. J. Beaver, Professor of Chemistry in Columbia University. The phenyl ethyl alcohol used in the research was generously supplied by Givaudan-Delawanna, Inc., through Mr. Edward Sagarin.

as a dual purpose (1) the study of olfaction and (2) the verification of the Weber-Fechner law by some method which would reveal increase of sensation with added stimulation. Four nine-step series of aqueous solutions of camphor were used as stimuli. The weakest concentration, i.e., the initial point in the first series, contained one part of camphor in 1,000,000 parts of water. The next series began with 1/100,000 (or 10/1,000,000), the third with 1/10,000, and the fourth with 1/1,000. In testing, the *S* sniffed the bottle of camphor and compared the sensation with that resulting from sniffing a bottle of distilled water. The authors conclude ". . . that the Fechner-Weber law is true in principle, in that stimulation must increase faster than sensation, but not in its mathematical formula" (**19**, p. 641).

Hermanides measured the discrimination thresholds for nine substances by using the Zwaardemaker olfactometer and found them to be very much in accord with those of Gamble. His results as reported by von Skramlik (**18**, p. 305) show relative limens varying from about 25 to 60 percent. Von Skramlik states, "We conclude from these facts that in the case of the sense of smell Weber's Law is not easy to prove and one must be satisfied with the values of Hermanides."

The three studies just mentioned were all concerned with establishing Weber's Law for the olfactory sense. Zigler and Holway (**24**) stated as their purpose an "experimental account . . . of the way in which differential sensitivity varies as a continuous function of some measurable aspect of the stimulus" (**24**, p. 372). Amount of olfactory substance was selected as the measurable aspect. They too used the Zwaardemaker olfactometer with India rubber as the stimulating substance and obtained data by the method of single stimuli with two response categories. Within the range of 10 to 400 olfacties, as the stimulus amount was increased, the Weber fraction was found to decrease from .99 to .17.

All of the experiments just described are subject to some criticism regarding the technique used for producing the stimulus. The erroneous assumptions underlying use of the Zwaardemaker olfactometer, and hence of the olfactie, have been discussed previously (**9**). There is need for measurement of olfactory sensitivity by carefully controlled means. The present experiment explores one possibility of achieving control over certain significant variables.

Experimental Procedure

Apparatus.—The goal in designing and constructing the olfactometric instrument used in this experiment was to be able to express each stimulus in terms of the number of odorous molecules contained in it. To meet this goal, two significant aspects of the stimulus had to be controlled, viz., temperature and volume. In order to regulate volume, a much modified form of Elsberg's blast injection technique (**5**) was used. Each stimulus was blown into *S*'s nostril under a pressure in known excess of atmospheric. Hence, it was necessary actually to control pressure and duration to achieve ultimate control over volume. In addition, the atmospheric pressure at the time of each experimental session had to be recorded and the vapor pressure of the stimulus substance at the working temperature had to be known.

Phenyl ethyl alcohol was chosen as the stimulus substance because it has no trigeminal effects (**6**), is pleasant and delicate, and does not fatigue over a long period of time.

A diagram of the main features of the apparatus is shown in Fig. 1. The course of the air through the system is from right to left. Tank 1, of nine-gallon capacity, is filled to any pressure under 20 pounds by a diaphragm compressor (Pump) with air from the surrounding room. Small quantities of air are then drawn from this tank as needed. Air flows into Tank 2 when the two stop-cocks between it and Tank 1 are opened. (The Exit Tube serves only to release excess air quickly from Tank 2 if desired for any reason. Ordinarily, the stop-cock in this line is kept closed.) The air stream passes first through an activated carbon filter to remove organic impurities and then through a gas washing bottle filled with the stimulus liquid. This bottle is equipped with a special fritted disk at the end of the inlet tube to break up the air current into very small bubbles, maximizing the contact between air and liquid.

In Tank 2, a duplicate of Tank 1, temperature and pressure controls are introduced. The tank

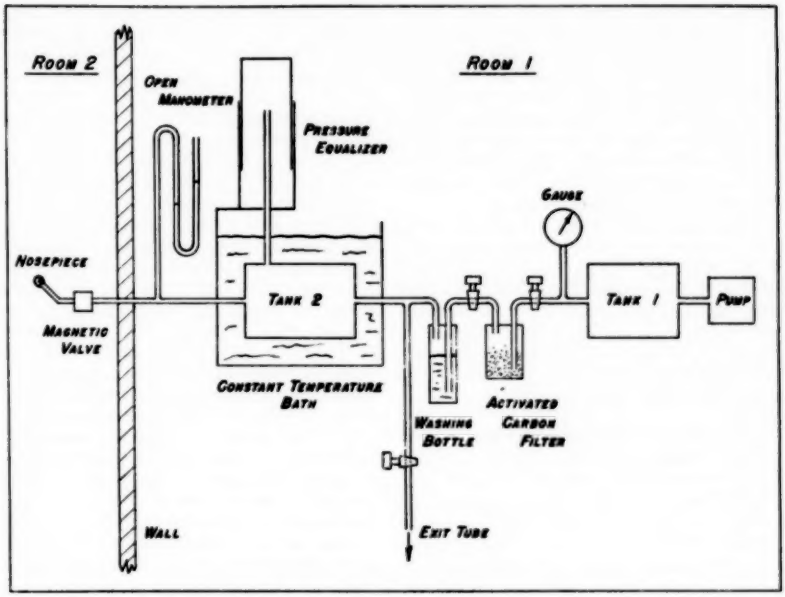


FIG. 1. Diagram of apparatus

rests in a constant-temperature bath where temperature is maintained at 20° ± .1° C. The essential parts of this bath—heating and cooling coils, stirrer, thermostat, and mercury relay—have not been included in the diagram. As the vapor rests in this tank, the mixture of air molecules and molecules of odorous material reaches an equilibrium according to the regulated temperature. Inside the tank is an evaporating dish holding the stimulus liquid. The vapor can be compressed at will in varying amounts by means of the metal unit labelled Pressure Equalizer in the diagram. This unit is similar in construction and operation to a wet spirometer or a gas dispensing tank. It is directly connected with the interior of Tank 2. The bottom cylinder of the unit is half-filled with distilled water which acts as a seal. The amount of compression within the tank varies according to the weight which *E* places upon the upper cylinder. When the system is opened to deliver a stimulus, the upper cylinder falls so that the pressure is evenly maintained throughout the duration of the stimulus. The pressure may be read on the open manometer in terms of mm. of mercury (mm. Hg) above atmospheric pressure. The manometer fluid is red gauge oil, with specific gravity 1/17 that of mercury. Thus, it is easy to measure the small changes in mercury units very accurately, since the red gauge oil moves 17 times as far as mercury.

All of the parts so far described are located in Room 1, entirely removed from *S*. In Room 2, just on the other side of the wall, are the magnetic valve which releases the stimulus, and the nosepiece. The distance from Tank 2 to the valve is 36 cm., an absolute minimum. The electrical control for the valve is a revolving disk powered by a Telechron constant-speed motor of one rpm. The disk, located in Room 1, also controls a warning light in front of *S* which goes on two sec. before the valve is opened. The duration of the stimulus and the interval between stimuli can be varied independently by adjusting the contacts on the disk. Throughout the present experiment, stimulus duration was kept constant at .50 sec. and the interval between stimuli was kept at 29.50 sec. Room 2 was also equipped with a telegraph key in circuit with a light in Room 1 for *S*'s response signals. Each *S* had his own nosepiece. They were specially constructed according to a uniform design to fit almost any nostril tightly. The bulb is an elongated oval of solid glass built around the same size tubing used throughout the apparatus. The angle of the nosepiece tube facilitates its insertion and directs the air upward toward the olfactory membrane.

All construction materials are odorless. The tanks are stainless steel and glass is used wherever possible in the remainder of the apparatus. Joints are effected by rubber sleeves. The connecting ends of glass tubing meet and the rubber merely holds them in place and makes the joint airtight. All rubber is pure gum, specially treated to remove the odor. The internal diameter of all glass tubing is six mm. Minimal lengths are used to facilitate cleaning.

Both rooms can be ventilated readily by turning on a blower system which is permanently installed in the experimental rooms. The system circulates fresh air taken from outside, from a heated source, or from a combination of the two. Room temperature was always higher than the temperature in Tank 2 to prevent condensation in the tubing from Tank 2 to *S*. It rarely fluctuated more than two degrees from 26.5° C. (80° F.).

Calibration and calculations.—The stimulus intensity—defined here as number of molecules—is varied by changing the stimulus volume. The volume, in turn, is varied through control of the pressure by means of the weight on the descending cylinder of the pressure equalizer. Calibration of the volumes was accomplished as follows. The descending cylinder was graduated in mm. It had previously been determined that a distance of 13 mm. was equivalent to 200 cc. For each of five different weights of the cylinder, the extent of fall in five sec. was measured. Ten such readings were taken for each weight. Each set of 10 readings was then averaged. The mean was translated into cc. and divided by 10 to give the volume for .50 sec. The resulting five points were plotted on a graph so that the volume for any intermediate weight could be read. For conversion of weight to pressure, all that was necessary was to read the manometer and convert to mercury units. The relationship between weight and volume is very slightly negatively accelerated; that between weight and pressure is rectilinear within our limits.

Conversion into molecular count was done by calculation. The first step was to determine the vapor pressure (VP) for the substance at 20° C. from measured values (17, p. 259) at higher temperatures. This value was calculated as .06 mm. Hg. To find the ratio of the odorous molecules to the molecules of air (r) the vapor pressure is divided by the total pressure (TP) in the system which is the sum of the atmospheric pressure and the excess pressure ($r = VP/TP$). To find the volume (V) that one gram molecular weight of the substance would occupy under these conditions of temperature and pressure, the standard volume (22,400 cc.) which obtains at 0° C. (273° K.) and 760 mm. Hg is corrected for the total pressure and 20° C. (293° K.) [$V = (760/TP)(293/273)(22{,}400)$]. To find the total number of molecules (N) in the delivered volume (v), the ratio of v/V is multiplied by Avogadro's number (6.02×10^{23}), which is the number of molecules in one gram molecular weight of any substance [$N = v/V(6.02 \times 10^{23})$]. To find the number of *odorous* molecules (M) in the delivered volume, N must be multiplied by r ($M = Nr$). It is obvious that M will vary somewhat as the atmospheric pressure varies. The actual extent of this variation as determined by calculating M for each weight on each day that it was used was never more than 1.5 percent. Since the variations were so slight, the modal calculated value for each weight has been used to represent the molecular count for that stimulus in all psychophysical computations. The amount of error to be expected due to the variability of the determinations of the various steps in the calculations is not greater than two percent. This maximum variability occurred in measuring the stimulus volume. The level of intensity represented is well above the average absolute threshold which is calculated to be approximately 10^4 molecules (14, pp. 79, 315–6).

Method.—Our own preliminary studies and all of the published reports emphasize the fact that differential sensitivity in olfaction is not easily measured by standard techniques. *S*'s task in a psychophysical experiment is inimical to long-established habits of noticing only the qualitative aspects of an olfactory stimulus in the relatively few cases where anything is noticed at all. As Gamble said, "Most persons are not used to smelling attentively" (7, p. 109). For this reason, as well as the strong tendency of the receptor to become fatigued in a short time as compared with the eye or the ear, it is desirable to make the experimental procedure and *S*'s task as simple and brief as possible. The duration of each stimulus was very brief with an interval between stimuli long enough to offset the effects of fatigue according to Elsberg (4).

The psychophysical method used on all *S*s was the modified form of the method of single stimuli described by Koester and Schoenfeld (12). Pfaffman (16) compared the standard methods of constant stimuli and single stimuli in gustation and found that the two procedures gave the same results. Gustation and olfaction are analogous modalities in regard to the difficulties of controlled stimulation. The modified form of single stimuli consists of presenting in any given series only one pair of stimuli arranged equidistantly about a hypothetical midpoint. *S* responds to each stimulus member of the pair with a plus or minus judgment, indicating more or less of the attribute being measured. In this experiment only a single pair was used on any one day. *S*s were instructed to make judgments

of the amount of odor only and to disregard the stimulus pressure. They were given at least three practice sessions to familiarize them with the nature of the odor and with the stimulus characteristics, and to enable them to build up criteria for judgment. They indicated a judgment of 'Weak' by pressing the key once, of 'Strong' by pressing it twice. A complete session consisted of 40 presentations, 20 of the weak stimulus and 20 of the strong. A session was preceded by three identified samples of each stimulus, weak and strong alternating, to assist *S* as much as possible in finding and maintaining criteria for judging. Each stimulus pair was used in two separate sessions, making a total of 80 judgments for the pair or 40 judgments per stimulus. The same nostril was used during all of each *S*'s sessions. The 40 stimuli in each session were presented in blocks of 10 with three to five min. rest-periods between blocks. (*S*: ES requested in addition one and a half to two min. rest between stimuli 5 and 6 in each block.) During the rest-periods, the room was ventilated and *S* went out into the hall for a change of air. A session lasted from 40 to 50 min. The stimuli varied in a random sequence with two restrictions imposed: (1) that no stimulus be repeated more than three times in succession and (2) that half of the blocks begin with a strong and half with a weak stimulus.

Four pairs of stimuli were used with one *S*, five pairs with one *S*, and six pairs with two *S*s. Since a pair was used in two different experimental sessions for each *S* some evidence bearing on day-to-day variability was obtained. The sequence of pairs through each *S*'s different sessions was approximately from the greatest difference down to the smallest difference and up again to the largest. In the case of three *S*s, the smallest difference was used in the final two sessions when it had become clear that the differences for which the data had already been gathered did not yield p values (percentage of Strong judgments) close enough to .50.

In addition to the odor series just described, a control series duplicating the experimental series except for lack of odor was run on one *S* (KP). The instructions here were to judge in terms of the stimulus pressure. (*S* reported transient odors of various types from time to time during these sessions but for consistent comparisons pressure was the only identifiable dimension of variation.) The purpose was to investigate the possibility that the functions obtained under the 'Odor' instructions were no different from those obtained under 'Pressure' instructions. If the two functions should be demonstrably different, it would be strong evidence fot the measurement of true odor discrimination. If they should be similar, however, it would nor necessarily support the conclusion that the supposed odor discriminations amounted merely to actual pressure discriminations. It would suggest that this might be the case but further experimentation would be necessary to bring out more conclusive behavioral evidence. It is possible, of course, that the two functions really do have the same shape even though the *S* is truly performing two separate tasks.

Subjects

Complete data were obtained from four *S*s, three women and one man (ES). The women are all graduate students in psychology who have had some training as *S*s in psychophysical experiments. KP had served extensively in a preliminary experiment carried out by the author two and a half years before. Neither GG nor CF had had any previous experience in olfactory experiments. ES had never served in a psychological experiment but had had years of experience in 'smelling attentively' since he is connected with a large aromatics company. He had no difficulty in adjusting to the demands of the experiment and exhibited a cooperation and devotion to duty rarely seen in non-professional *S*s.

All experimental sessions were completed between Feb. 2 and Mar. 4, 1948. During that time, only one *S* (CF) had a cold. The infection was of brief duration and experimentation was resumed 10 days after onset of the symptoms. There was no evidence—introspective or behavioral—that any impairment in acuity still remained at that time. One other *S* (GG) had sinus difficulties on the very few warm wet days within this interval. Therefore, no sessions were held for heri n such weather.

Results

The raw data consist of 40 responses to each stimulus, 20 collected on one day and 20 on another day. A single response is either 'Strong' or 'Weak.' Each *S*'s data have been treated separately. Table I shows the percent of Strong responses to the single members of each stimulus pair on each day that it was used and for both days put together. Table II shows the mean and standard deviation for each pair derived from the p values for the pair. For example, *S*: GG when presented with stimuli 1.42 and 2.06×10^{17} molecules for the first time

gave 40% Strong responses to 1.42 and 95% Strong responses to 2.06; the second time gave 5% Strong responses to 1.42 and 60% to 2.06. Combining the data for these two days gives 22.5% Strong responses to 1.42 and 77.5% to 2.06 (cf. Table I). Using these p values, the mean for the pair, 1.42 and 2.06 $\times$ 10^{17} molecules, is found to be 1.57 $\times$ 10^{17} molecules on Day 1, 1.97 on Day 2, and 1.74 when the data from the two days are combined. The standard deviation based on the p values is .34 $\times$ 10^{17} molecules on Day 1, .34 on Day 2, and .42 for the combined data. The mean obtained from a set of judgments tells us what hypothetical stimulus magnitude served as the midpoint of the distribution of judgments. It is equivalent, then, to the PSE. This figure can be compared with the actual midpoint of the pair of stimuli to give an indication of the direction of the constant error. The standard deviation obtained from the judgments is a measure of S's sensitivity. It tells us the extent to which the

TABLE I

P Values for 'Strong' Responses to the Members of Each Stimulus Pair on Each of Two Days Presented and for Both Days Combined

N per day equals 40. Bracketed p values are actually .00 and 1.00 but were changed for computation. Stimulus unit is 10^{17} molecules; e.g., the first pair of stimuli listed for Subject: GG are 1.42 $\times$ 10^{17} molecules and 2.06 $\times$ 10^{17} molecules.

Stimuli	p values			Stimuli	p values		
	Day 1	Day 2	Days 1 & 2		Day 1	Day 2	Days 1 & 2
Subject: GG				Subject: CF			
1.42	.40	.05	.225	1.68	.30	.20	.25
2.06	.95	.60	.775	1.85	.75	.60	.675
1.48	.20	.15	.175	Subject: KP			
2.02	.85	.55	.70	1.42	.05	(.01)	.03
1.54	.35	.20	.275	2.06	.75	(.99)	.87
1.98	.65	.90	.775	1.54	.20	.10	.15
1.60	.20	.15	.175	1.98	.70	.90	.80
1.93	.70	.70	.70	1.63	.50	.20	.35
1.68	.35	.35	.35	1.91	.65	.80	.725
1.85	.60	.60	.60	1.68	.45	.25	.35
1.74	.35	.50	.425	1.85	.80	.85	.825
1.79	.75	.45	.60	1.74	.60	.50	.55
Subject: CF				1.79	.40	.70	.55
1.30	(.01)	(.01)	(.01)	Subject: ES			
2.13	.90	.90	.90	1.42	.20	.05	.125
1.42	.20	(.01)	.105	2.06	.95	(.99)	.97
2.06	.60	.90	.75	1.54	.30	.25	.275
1.48	.10	(.01)	.055	1.98	.85	.95	.90
2.02	.70	.85	.775	1.66	.50	.35	.425
1.54	.30	.35	.325	1.88	.85	.55	.70
1.98	.60	.85	.725	1.71	.60	.40	.50
1.60	.20	.20	.20	1.82	.70	.65	.675
1.93	.65	.80	.725				

TABLE II

MEAN AND STANDARD DEVIATION FOR EACH STIMULUS PAIR ON EACH OF TWO DAYS PRESENTED AND FOR BOTH DAYS COMBINED

Unit is 10^{17} molecules

Stimulus Pair	Mean			Standard Deviation			Stimulus Pair	Mean			Standard Deviation		
	Day 1	Day 2	Days 1 & 2	Day 1	Day 2	Days 1 & 2		Day 1	Day 2	Days 1 & 2	Day 1	Day 2	Days 1 & 2
Subject: GG							Subject: KP						
1.42, 2.06	1.51	1.97[a]	1.74	.34	.34	.42	1.42, 2.06	1.87	1.74[a]	1.82	.28	.14[a]	.21
1.48, 2.02	1.72	1.96[a]	1.83	.29	.46[a]	.37	1.54, 1.98	1.81	1.76	1.78	.32	.17[a]	.23
1.54, 1.98	1.76	1.71	1.73	.57	.21[a]	.33	1.63, 1.91	1.63	1.77	1.74	.73	.17[a]	.28
1.60, 1.93	1.80	1.82	1.81	.24	.21	.23	1.68, 1.85	1.70	1.75	1.74	.18	.10[a]	.13
1.68, 1.85	1.78	1.78	1.78	.27	.27	.27	1.74, 1.79	—[d]	1.74	—[c]	—[d]	.10	—[c]
1.74, 1.79	1.76	—[d]	1.76	.05	—[d]	.11	All combined			1.77			.22
All combined			1.78			.31							
							Subject: ES						
Subject: CF							1.42, 2.06	1.64	1.69	1.66	.26	.16[a]	.21
1.30, 2.13	1.84	1.84	1.84	.23	.23	.23	1.54, 1.98	1.69	1.67	1.68	.28	.19[a]	.23
1.42, 2.06	1.91	1.83	1.84	.58	.18[a]	.33	1.66, 1.88	1.66	1.83[b]	1.72	.21	.43[a]	.31
1.48, 2.02	1.86	1.85	1.85	.30	.16[a]	.23	1.71, 1.82	—[c]	1.75	1.71	—[c]	.17	.24
1.54, 1.98	1.84	1.66	1.73	.57	.31[a]	.42	All combined			1.69			.25
1.60, 1.93	1.83	1.77	1.79	.27	.20	.23							
1.68, 1.85	1.75	1.81	1.78	.14	.16	.15							
All combined			1.80			.27							

[a] Value of t indicates significant difference from Day 1 at .01 level or better.
[b] Value of t indicates significant difference from Day 1 at .05 level.
[c] Cannot be computed, since both p values exceed .50.
[d] Cannot be computed, since p values are reversed.

actual midpoint of the pair (serving as an hypothetical standard) must be varied in intensity in order to make the variation detectable in 84 percent of the cases. M and SD were computed by using simultaneous equations assuming a normal ogive as suggested by Woodworth (**22**, p. 408–9). When both p values exceed .50 or when the weaker stimulus receives a p value greater than .50 while the stronger stimulus is given a value less than .50, no M or SD can be computed. Values of .00 or 1.00 cannot be used for computation either. Whenever these values appeared, they were changed to .01 and .99 so that they could be used. Results for the two days on which a given pair was used are shown separately as well as the values computed from the combined data of the two presentations.

All of the differences between means and between standard deviations from Day 1 to Day 2 were tested for significance by the t-test. The SD's were considered as uncorrelated and the ordinary formulas applied, using N as the number of times each stimulus was presented.[2] The possible significance of any of these differences is of interest in regard to the day-to-day variability of the S but does not invalidate combination of the data. Each of the two means for a given stimulus pair is an independent estimate of the mean of the same popula-

[2] It is recognized that there is a problem here in determining the reliability of the SD, since the latter is not the SD of the judgments making up the p values for two particular stimuli but rather is the SD of an entire distribution of p values which is inferred from the two p values available. Precedent for the present procedure can be found in Woodworth (**22**, p. 414) and Guilford (**8**, p. 179).

tion, viz., *S*'s subjective midpoints. The two estimates may differ significantly, indicating that the subjective midpoint has fluctuated. The same is true of the *SD*'s or sensitivity estimates. By combining the two sets of

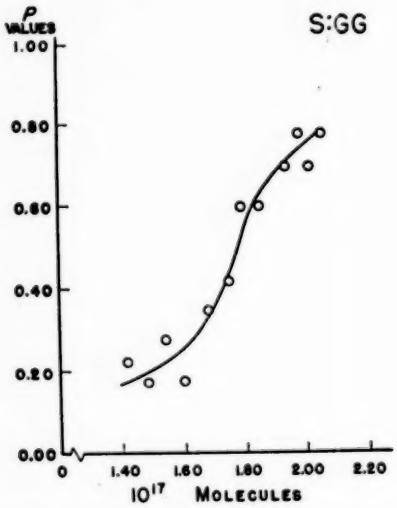


FIG. 2. Percent of strong responses to different numbers of odorous molecules for *S:* GG

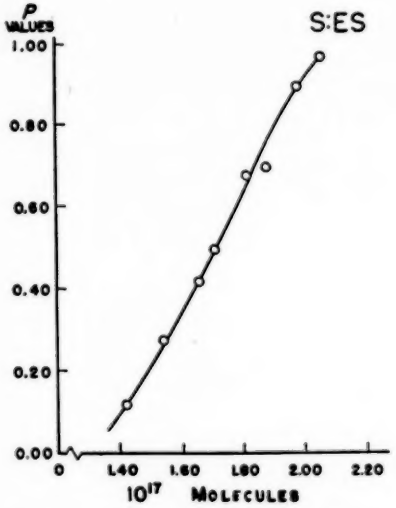


FIG. 4. Percent of strong responses to different numbers of odorous molecules for *S:* ES

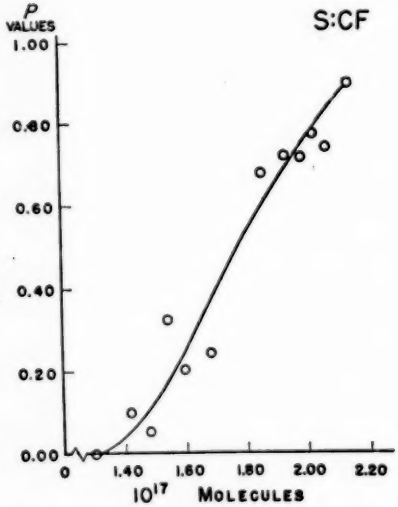


FIG. 3. Percent of strong responses to different numbers of odorous molecules for *S:* CF

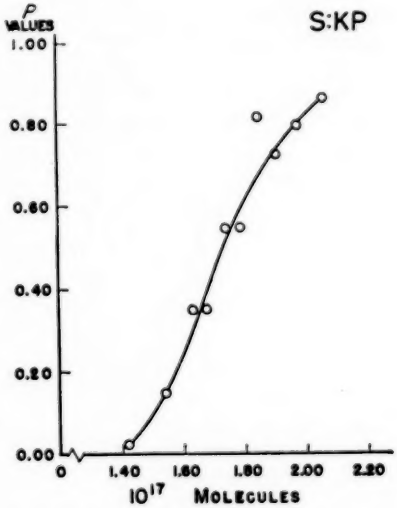


FIG. 5. Percent of strong responses to different numbers of odorous molecules for *S:* KP

TABLE III

WEBER FRACTIONS FOR EACH STIMULUS PAIR ON EACH OF TWO DAYS PRESENTED AND FOR BOTH DAYS COMBINED

SD equals Standard Deviation; *S* equals actual midpoint of stimulus pair serving as hypothetical standard. Stimulus unit is 10^{17} molecules.

Stimulus Pair	*SD/S*			Stimulus Pair	*SD/S*		
	Day 1	Day 2	Days 1 & 2		Day 1	Day 2	Days 1 & 2
Subject: GG				Subject: KP			
1.42, 2.06	.20	.20	.24	1.42, 2.06	.16	.08	.12
1.48, 2.02	.17	.26	.21	1.54, 1.98	.18	.10	.13
1.54, 1.98	.32	.12	.19	1.63, 1.91	.41	.10	.16
1.60, 1.93	.14	.12	.13	1.68, 1.85	.10	.06	.07
1.68, 1.85	.15	.15	.15	1.74, 1.79	—[a]	.06	—[a]
1.74, 1.79	.03	—[a]	.06	All combined			.13
All combined			.18				
				Subject: ES			
Subject: CF				1.42, 2.06	.15	.09	.12
1.30, 2.13	.13	.13	.13	1.54, 1.98	.16	.11	.13
1.42, 2.06	.33	.10	.19	1.66, 1.88	.12	.24	.18
1.48, 2.02	.17	.09	.13	1.71, 1.82	—[a]	.10	.14
1.54, 1.98	.32	.18	.24	All combined			.14
1.60, 1.93	.15	.11	.13				
1.68, 1.85	.08	.09	.09				
All combined			.15				

[a] Cannot be computed since *SD* is unobtainable.

data on which the two estimates are based, the result is a better over-all estimate of *S*'s experience.

Figs. 2–5 contain the smoothed discrimination functions of the four *S*s. The plotted points are the percent Strong responses to each stimulus, based on the total 40 judgments. The smoothed curves were drawn by inspection.

The Weber fractions are shown in Table III. These values show the relationship between the measure of sensitivity, i.e., the *SD*, and the standard, i.e., the actual midpoint of the pair. A single Weber fraction based on the *S*'s entire performance on all stimuli was computed by combining all of the *SD*'s and dividing the result by the average midpoint.

For the combined data of all *S*s on all stimulus pairs, the mean is 1.77 × 10^{17} molecules, the standard deviation is .27, and the Weber fraction is .15.[3]

In order to test the significance of the CE's, the procedure was followed of computing the E% for each day's judgment and for the two days combined. The E% is a measure of the percent of excess use of one response category as compared with the other. Thus, it can fluctuate in a plus or minus direction about zero, depending on which category is used more often. Excess use of the Strong category means that the subjective midpoint is below the actual midpoint and hence represents a minus CE. Excess use of the Weak category by the same reasoning indicates a plus CE. A *t*-test can be applied to ascertain the significance of the difference be-

[3] For the formula used to combine *SD*'s see Walker (20, p. 145) or Peters and Van Voorhis (15, p. 81).

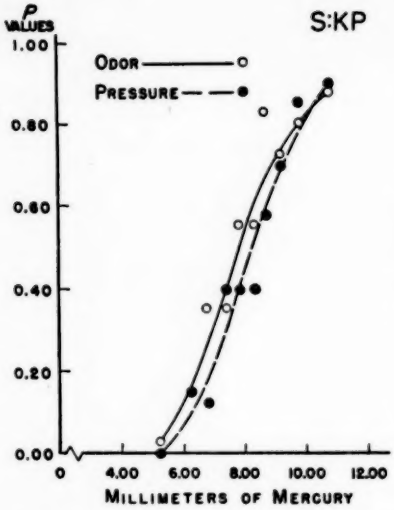


FIG. 6. Percent of strong responses to different numbers of odorous molecules and to different odorless pressures for *S*: KP. $\chi^2 = 5.7186$; $.70 > P > .50$. (Odor intensities are expressed in pressure units to make the two curves comparable.)

tween the obtained E% and zero. The rationale and the formulas involved have been discussed by Koester (11, p. 63–4). Only five percent of all 63 E%'s achieved significance at the .05 level or better, one plus CE and two minus CE's.

The results of the control series are shown in Fig. 6. The two functions, one obtained under 'Odor' instructions and one under 'Pressure' instructions, are plotted on the same axes. The abscissa values are in pressure units since pressure is the stimulus characteristic which the two conditions have in common. The difference between the two distributions when expressed as frequencies rather than percentages was tested for significance by chi-square. The two extreme stimuli were omitted because the frequencies for the weak member of that pair were too small to be handled. The hypothesis being tested is that the two distributions do not differ.[4]

DISCUSSION

Inasmuch as no work had been done on relative discrimination using a blast injection technique, it was not known whether *S*s could perform the task required without a prolonged training period. They all reported that their judgments were based on the strength of the odor and that the effects of the pressure were negligible at all differences except the extremes. However, the control series on one *S* failed to supply behavioral substantiation for the introspections. As discussed above in the section on Method, the control series does not constitute a conclusive check on the hypothesis that the *S*s were actually making pressure discriminations throughout. This hypothesis would be susceptible to a more conclusive test by running another control series in which odor judgments were given while pressure was held constant. If the function for odor alone did not differ from the functions for pressure alone and for odor plus pressure, it could safely be assumed that the two types of discrimination show the same distribution of judgments and the *S*s' reports could safely be accepted. If the present control had revealed a significant difference between the two distributions, no further test would be necessary. If pressure had been of assis-

[4] The formula used was:

$$\chi^2 = \frac{N^2}{AB}\left[\Sigma\frac{a_i^2}{n_i} - \frac{A^2}{N}\right],$$

where a is the frequency of Strong responses for each stimulus under Odor instructions, A is the sum of the a's, B is the sum of the frequencies of Strong responses to the individual stimuli under Pressure instructions (i.e., the sum of the b's), n is the frequency of Strong responses to each stimulus under both conditions (i.e., $a + b$), and N is the sum of the n's (i.e., $A + B$) (21).

tance in making the judgments, the best discrimination as measured by the *SD* would be expected at the largest differences where pressure distinctions were relatively clear-cut. No *S* showed such an effect, however. In the absence of conclusive evidence favoring pressure discriminations and in the presence of strong introspections favoring odor discriminations, the data are assumed to be expressions of olfactory sensitivity although the other possibility is recognized.

Jerome (10) reported similar results in his paper on absolute thresholds in which the stimulus unit was pressure in mm. Hg. His *S*s sometimes signalled sensation of odor when the blast was delivered from an odorless bottle. The problem is confused by Adrian and Ludwig's (2) report that potentials in the olfactory nerve can be picked up following odorless pressure on the olfactory receptor. The potentials recorded were from a catfish preparation. No one has reported any results from a specimen with an olfactory system more similar to a human's, although Adrian (1) has described the effect of stimulating the receptor mechanically on the electrical activity of the olfactory bulb and the pyriform area of the hedgehog.

Inspection of Figs. 2–6 reveals consistent ogival tendencies but no clear-cut realization of such a shape. Several factors may be considered as contributing to the irregularities. (1) The number of judgments was only 40 for each point due to the time required to obtain each one, so that none of the p values has been established as reliably as if 100 judgments had been given per point. (2) Regardless of the fact that the *S*s were able to perform the task, it was still a newly-acquired proficiency for at least three of them. As is customary with new skills, performance fluctuates somewhat until a certain degree of automaticity has been acquired. This principle applied to sensory discrimination as well where automaticity plays a role in insuring uniformity of criteria. In the light of previously published accounts of the difficulties and inconsistencies encountered in olfactory experimentation, the orderliness of our data becomes the outstanding characteristic. (3) The *S*'s sensitivity, as measured by the pattern of responses to the same stimulus pair, varied significantly. Whether this change was merely the result of fluctuating performance due to shifting criteria as discussed above, or whether it was due to some variation in the condition of the receptor cannot be determined definitely. In the case of the former, this type of variability would be expected to disappear as familiarity with the task increased. Within the limits of the experiment, there is no sign of such a tendency. ES, with ample experience at attending to odor intensities, showed as many significant fluctuations in sensitivity as the other *S*s. The latter possibility has more evidence in its favor. In a city such as New York there are climatological conditions, as well as the heavy soot content of the air, which are known to cause nasal irritation. In some cases the result is a greater mucous secretion and a consequent blocking, partial or complete, of the passage to the olfactory area. In other cases there is no apparent blocking but rather *S* reports simply that he "just can't smell much today." Perfumers claim that New York, San Francisco, and New Orleans are the worst cities in the U. S. for proper functioning of the olfactory sense. Each of our *S*s reported at least once that the whole level of acuity seemed to have shifted so that both stimuli gave the impression of being much weaker or stronger than the general intensity at which he had been working. We may assume, then, that even if the number of judgments were increased and further proficiency was acquired, variations in sensitivity would remain to a certain extent to cause some irregularities in data based on more than one day's work.

It is of interest to compare the sensitivity as measured by this technique with the reports of other investigators who used different appara-

tus and different methods. The Weber fractions in Table III are expressions of each *S*'s sensitivity in relation to the intensity level. The largest Weber fraction for any *S* is .41 and the smallest is .03. The total number computed (excluding those for all stimuli combined) is 59. Forty-eight are less than .20. Gamble reported that the mode for relative thresholds in her experiment was .33 and the range was from less than .17 to more than .50. Hermanides reported no values less than .24. Zigler and Holway reported relative thresholds less than .20 for only the highest of six intensity levels used (400 olfacties). Using the over-all estimate based on our four *S*s, it appears that olfaction is not as sensitive in the detection of differences as are vision and audition, although it is not as far behind the latter as might be expected. Since the need for such acuity is rarely, if ever, forced upon us, we are apparently unaware of our capabilities.

The results of this experiment show keener differential sensitivity than do the results of previous researches when the Weber fractions are considered. This is especially true when it is realized that the measure of sensitivity used here as the numerator of the fraction, i.e., the *SD* or 84% threshold, is larger than the measure used by Gamble which was the 50% threshold obtained by the method of just noticeable differences. Zigler and Holway also used the 84% threshold. The difference might be explained by postulating a lower relative threshold for phenyl ethyl alcohol than for the substances used by all of the other workers. This hypothesis is not tested by the present experiment. It would be necessary to use these same *S*s in further experimentation with some of the substances used by Gamble and others. The Zigler and Holway results suggest that the difference might be due to the level of intensity at which we worked. It is impossible, however, to compare our level with theirs, since the olfactie cannot be translated into molecular count. Hence, to test this hypothesis, it would be necessary to collect further data from the same *S*s at other levels. This explanation is not supported by coments from some *S*s that the weakest stimulus occasionally was hardly noticeable, which suggests that our stimulus range was not on a high intensity level. Rather, there are two reasons for believing that the present measurements furnish a more accurate estimate of olfactory sensitivity. First, the stimulus material did not excite the trigeminal nerve endings as did many of Gamble's and Hermanides' materials. That means that *S*'s response was more apt to be based upon olfactory sensations alone except for the possible influence of the pressure factor. Second, the experimental conditions such as stimulus volume, stimulus duration, and temperature of both the substance and the surroundings, were controlled. Also, stimulation time was minimal and *S* was not allowed to become adapted to any measurable extent. The variability caused by fluctuations in these extraneous conditions has been eliminated, leaving the variability caused by the *S*'s own changes in sensitivity to a constant stimulus intensity. It is felt, therefore, that the measurements reported here are a better approximation of acuity in this modality than those previously reported.

It might be expected that *S*: ES, having had a great deal of previous training in olfactory discrimination, would show keener sensitivity than the other *S*s. His data are the most consistent of all the *S*s and so his thresholds do not differ as widely

among themselves. Generally speaking, he is not any more acute than the others. It seems very likely, however, that his absolute sensitivity is keener than that of the others. This is a tentative conclusion based on casual observations only.

Such a small percentage of the constant errors achieved significance that we can assume that, in general, no subjective displacement of the midpoint occurred. The mean based on the combined data of all *S*s is equal to the actual midpoint of the stimulus series. The strongest hint of such a displacement is found in the data for *S:* ES, where seven out of eight CE's were in the minus direction indicating preponderance of Strong judgments. It is possible that, although almost none of his CE's are significant in regard to magnitude, the consistency of the trend is significant. Support for the latter can be found in the introspections. Because of his previous experience with very small amounts of odor, all of the amounts used here seemed large. His tendency was to use the Strong category more often.

The author feels that the present results provide more adequate measurements of differential sensitivity in olfaction than have been available heretofore. They may prove to be of assistance, also, in establishing the true nature of the stimulus. There is reason to believe that mere number of molecules is not the complete description of the olfactory stimulus. The recent theory of Beck and Miles (**3, 13**) suggests that odor results from absorption by the odorous substance of infrared waves emitted by the olfactory area with subsequent transformation into nervous impulses initiated by the change in thermal equilibrium. Different odors would then be the result of different patterns of absorption. Since infrared absorption spectra are actually expressions of intra-molecular activity it would seem important for any theory based on molecular phenomena to have available information on sensitivity stated in molecular terms.

Summary and Conclusions

The present experiment was conducted for the purpose of measuring differential sensitivity in olfaction by means of a modification of the blast injection technique. By introducing controls over the temperature, pressure, and volume of the stimulus, and by using an odorous compound with a calculable vapor pressure, it was possible to express the stimulus intensity in terms of the number of odorous molecules. Four *S*s were tested for discrimination of amount of odor by the modified method of single stimuli. Complete psychophysical functions were drawn for all; and means, standard deviations, Weber fractions, and constant errors were computed for each pair of stimuli as presented. Since each pair was presented on two separate days to every *S*, it was possible to determine the significance of the day-to-day variability by the *t*-test. In the odor series the stimulus varied not only in amount of odor but also in pressure. Consequently, a control series was run on one *S* in which amount of pressure was to be discriminated, the odorous material being removed. A chi-square test was applied to the difference between the two distributions of judgments obtained under the two conditions and instructions.

On the basis of the results, the following conclusions can be drawn.

1. The Weber fraction in olfaction, using phenyl ethyl alcohol as the test substance, is about .15. This degree of sensitivity is considerably lower

than any previously reported by other investigators using techniques with fewer controls over significant stimulus characteristics and using other substances.

2. Day-to-day fluctuations in sensitivity as measured by the standard deviations, are generally significant. Most differences between means of different days are not significant, however.

3. There is neither a consistent nor a significant constant error for three *Ss*. One *S* shows a consistent but insignificant minus deviation.

4. One *S* who had the benefit of considerable previous experience in making olfactory discriminations does not show a smaller Weber fraction than the three inexperienced *Ss*.

5. No significant difference between the distribution of odor judgments and that of pressure judgments could be demonstrated. The introspective data show that the *Ss* felt that they were disregarding the pressure while judging odor. Satisfactory demonstration of this impression at the behavioral level remains to be provided. In the meantime, since the blast injection technique is in fairly frequent use, the author advises caution in applying it to differential measurements until a further report has been made.

6. The results reported in this paper are thought to be more consistent with developing theory concerning the fundamental nature of olfaction than earlier reports since a molecular unit has been used to express stimulus intensity.

(Manuscript received April 9, 1948)

References

1. Adrian, E. D. Olfactory reactions in the brain of the hedgehog. *J. Physiol.*, 1942, **100**, 459–473.
2. Adrian, E. D., & Ludwig, C. Nervous discharges from the olfactory organs of fish. *J. Physiol.*, 1938, **94**, 441–460.
3. Beck, L. H., & Miles, W. R. Some theoretical and experimental relationships between infrared absorption and olfaction. *Science*, 1947, **106**, 511. (Abstract)
4. Elsberg, C. A. The sense of smell. VIII. Olfactory fatigue. *Bull. neurol. Inst. N. Y.*, 1935–36, **4**, 479–495.
5. Elsberg, C. A., & Levy, I. The sense of smell. I. A new and simple method of quantitative olfactometry. *Bull. neurol. Inst. N. Y.*, 1935–36, **4**, 5–19.
6. Elsberg, C. A., Levy, I., & Brewer, E. D. The sense of smell. VI. The trigeminal effects of odorous substances. *Bull. neurol. Inst. N. Y.*, 1935–36, **4**, 270–285.
7. Gamble, Eleanor A. McC. The applicability of Weber's law to smell. *Amer. J. Psychol.*, 1898, **10**, 82–142.
8. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
9. Gundlach, R. H., & Kenway, Geneva. A method for the determination of olfactory thresholds in humans. *J. exp. Psychol.*, 1939, **24**, 192–201.
10. Jerome, E. A. Olfactory thresholds measured in terms of stimulus pressure and volume. *Arch. Psychol.*, N. Y., 1942, **39**, No. 274.
11. Koester, T. The time error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level. *Arch. Psychol.*, N. Y., 1945, **41**, No. 297.
12. Koester, T., & Schoenfeld, W. N. Modified methods of constant and single stimuli in pitch discrimination. *Amer. J. Psychol.*, 1945, **58**, 540–544.
13. Miles, W. R., & Beck, L. H. Infrared absorption in field studies of olfaction in bees. *Science*, 1947, **106**, 512. (Abstract)
14. Moncrieff, R. W. *The chemical senses.* New York: Wiley, 1946.
15. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
16. Pfaffman, C. An experimental comparison of the method of single stimuli and the method of constant stimuli in gustation. *Amer. J. Psychol.*, 1935, **47**, 470–476.
17. Rechenberg, C. v. *Einfache und fraktionierte Destillation in Theorie und Praxis.* Miltitz bei Leipzig: Schimmel, 1923.
18. Skramlik, E. v. *Handbuch der Physiologie der niederen Sinne.* Band I. *Die Physiologie des Geruchs- und Geschmacks-sinnes.* Leipzig: Thieme, 1926.

19. TOULOUSE, E., & VASCHIDE, M. Note sur un nouveau moyen de vérifier la loi de Weber-Fechner sur la rapport de la sensation à l'excitation et sur la vérification de cette loi par la mesure de l'odorat, au moyen des solutions décimales. *C. r. Soc. biol.*, 1899, **51**, 640–642.

20. WALKER, HELEN M. *Elementary statistical methods*. New York: Henry Holt, 1943.

21. WALKER, HELEN M. Unpublished text material for course in statistical inference.

22. WENZEL, BERNICE M. Techniques in olfactometry: A critical review of the last one hundred years. *Psychol. Bull.*, 1948, **45**, 231–247.

23. WOODWORTH, R. S. *Experimental psychology*. New York: Henry Holt, 1938.

24. ZIGLER, M. J., & HOLWAY, A. H. Differential sensitivity as determined by amount of olfactory substance. *J. gen. Psychol.*, 1935, **12**, 372–382.

25. ZWAARDEMAKER, H. *L'Odorat*. Paris: Doin, 1925.

CONCEPT FORMATION AND EMERGENCE OF CONTRADICTORY RELATIONS *

BY JAMES CANNON DIXON
University of North Carolina

Introduction

Without a common recognition of the principles of logical relation there could be no basis for consistency and no distinction between what is possible and impossible. But could it appear self-evident to a being modeled after a non-dimensional mathematical point that *up* is the oppostie of *down?* If we consider relations as aspects or qualities which can be predicated only of two or more things taken together, the answer would plainly be in the negative, since the terms employed would have no meaning to such a being, at least according to our time- and space-bound way of thinking of such a being.

Spearman (5) made the eduction of relations a basic postulate among his noegenetic principles. This principle he states as follows: "The mentally presenting of any two or more characters tends to evoke immediately a knowing of relation between them." Expressed otherwise, Spearman would say that upon understanding what two concepts mean, e.g., *possible* and *impossible*, a person can immediately educe a relation; he will at least be able to say that the meanings have the relation of difference. Here, relations inhere in the very nature of the concept. Relations have come to have another meaning, however, that of connection or association between ideas (concepts). This seems to be the meaning given by Terman who took over the Absurdities of Binet in his Stanford Revisions (6, 7) as a measure of sensitivity to contradictory relations. He notes that, in the immature or stupid person, ". . . ideas do not cross-light each other, but remain relatively isolated. Hence, the most absurd contradictions are swallowed . . . without arousing the protest of the critical faculty." The implication seems to be that there can be an idea, *up*, and another idea, *down*, each of which remains relatively isolated without recognition of the relation between the two terms.

These two definitions of relations lead to two different sets of predictions. The former finds the ability to perceive relations whenever two concepts which are presented in relation are known. The latter finds this ability to be a later stage of development.

Problem

The notion of relations as inhering in the concepts presented, and the other view of relations as later learned connections can be set in problem form as follows: Does the child who demonstrates a functional use of the concepts *big* and *little* recognize the contradiction when these concepts are used in a contradictory relation, or does this behavior emerge as a later stage of development?

Procedure

Technique.—The contradictory situation used was the ring-segments illusion. The child is asked to pick up the 'big' segment, then position is interchanged and he is asked to pick up the 'little' segment. Since the lower segment always appears larger, interchange results in the previ-

* This report is an adaptation from a part of the doctoral research carried out under the direction of Dr. John F. Dashiell, University of North Carolina.

ously 'larger' now appearing smaller. In this way the child's present perception contradicts his immediately past perception and expressed judgment. Reaction to the contradictory relation posed by the illusion is inferred from reaction to the situation as to an unexpected event. To determine the child's ability to handle the concepts, the illusion was preceded by five sets of geometrical figures of different sizes and the child was required to pick up the big and little member at request as in the ring-segments illusion which constitutes a part of the series.

Subjects.—Since a random sampling of preschool children is infeasible, the children were obtained from nursery schools. Three schools in a university town, restricted largely to professors' and students' children, constitute what will be called the 'high occupational group.' Two schools in a nearby industrial city, catering mostly to the children of working mothers, will be referred to as the 'low occupational group.' A total of 143 children were tested. Occupational group, age and sex distribution of the children are shown in Table I.

TABLE I

OCCUPATIONAL GROUP, AGE AND SEX DISTRIBUTION OF CHILDREN

Age Group	Number of Children				
	High Group		Low Group		Total
	M	F	M	F	
2 to 3	3	6	4	7	20
3 to 4	8	9	9	6	32
4 to 5	12	12	13	15	52
5 to 6	9	9	8	13	39
Total	32	36	34	41	143

Materials.—The pairs of geometrical figures used in the experiment are represented in Fig. 1. They were made of six mm. thick cardboard for ease in handling. Shape, absolute size, order of arrangement, order of request, color of above and below members, as well as color of big and little members of a pair were varied systematically so as to leave no consistent basis for correct judgment other than relative size. (Since difficulty of size discrimination may prevent the child from demonstrating a reaction to the ring-segments, this possibility was controlled by designing three of the six pairs of figures to show slightly less apparent size discrepancy than the illusion. The rank order of discrepancy as judged from number of errors made was, from fewest errors to most: star, restangle, ring-segment, circle, square, triangle.)

Experimental procedure.—The child sat at a table on which were placed the first set of figures, the stars, one above the other with the larger nearer the child. Instructions were: "See the pretty big star and the pretty little star! We will play Big-and-Little with the pretty stars. Now this is the way we play. In this game I say, 'give me the *big* star' and I want to see if you can give me the *big* star. Then I say, 'give me the *little* star' and I want to see if you can give me the *little* star. Now, let's see if you know how to play Big-and-Little. Give me the *big* one." Then position is reversed and the experimenter says, "Give me the *little* one." (Incorrect responses are not corrected since this would introduce a variable learning factor.) The ring-segments are then presented in exactly the same manner as the other figures. After reversal, reaction to the illusion is rated independently by the experimenter and an observer. Then the test is repeated with the following variation. With the ring-segments in the original position, the experimenter says, "See, the *red* one is the *big* one and the *blue* one is the *little* one, isn't it? That's right! The *red* one is the *big* one and the *blue* one is the *little* one. Now give me the *little* one." After the child picks up the blue segment, it is replaced momentarily in the same position and the experimenter says, "See, the *blue* one is the *little* one, isn't it? Now [while reversing] give me the *big* one." Reaction is again rated.

Scoring.—On the concept test each correct choice between pairs of figures is scored one point. A total of 12 choices are made so the child's score may range from zero to 12. The actual scores (Table II) range from 3 to 12. (If the second choices after reversal on the two presentations of the ring-segments were counted, there would be 14 choices. These two choices do not constitute a part of the concept score, however, since after the reversal choice is complicated by conflict between previous judgment and present perception. Thus, these two choices are not comparable to the other 12.) A score of 7 is chance in the sense that if the child picked up the figure nearer him each time, he would be credited with 7 'correct' choices. Choice on any basis other than relative size would result in a score of 7 or less.

In scoring the test for reaction to contradictory relations a three point rating scale was used. The child's score is (+), (?), or (−) depending on whether he was rated as definitely, questionably, or not showing a reaction to the contradictory relation. Where there was disagreement between rating of the experimenter and observer, score was (?). Criterion for reaction to the illusion was behavioral expression (surprise, con-

fusion, hesitation, spontaneous verbalization, etc.) in picking up the ring-segment after reversal which was not shown with the other figures.

Reliability.—Observer agreement with experimenter was 92 percent. Reliability of the concept test was determined by using error scores on each figure and applying Kuder-Richardson formula 14, a measure of internal consistency (2). Reliability coefficient was +0.95. Reliability of the test for reaction to contradictory relations was determined by correlating rating on first presentation of the illusion with rating on the second. Those scoring (?) were combined with those scoring (+) in order to compute a phi coefficient using fourfold tables. (A null hypothesis that the (?) scores represented a random sampling of the (+) scores considered as a population, was found tenable.) The phi coefficient, corrected by Spearman-Brown prophecy formula, was +0.83, and is a measure of the extent to which the children scored either (+) or (−) on both presentations. (It will be noted that this is an underestimate, since the conditions for first and second presentations were somewhat different.)

Other indications of reliability may be summarily noted. The relation between concept test score and age, using an epsilon coefficient of correlation (since the relation found was curvilinear), was +0.61 with a standard error of 0.05. The high occupational group scored significantly higher than the low (*C.R.* = 3.327) but no significant sex difference was found. The epsilon coefficient computed from contradiction test scores and age was +0.57 with a standard error or 0.05. Again occupational group difference was significant and sex difference was not. It may be concluded, then, that increasing scores on both tests are strongly related to increasing age and that the brighter children (high occupational group) score significantly higher than the less bright (low occupational group), while sex was not an effective variable.

Findings

The relation found between score on concept test and rating on test for reaction to contradictory relations will indicate the answer to the problem of this investigation: Do the children who score high on the concept test tend strongly to be the ones who are rated as positively showing a reaction to the ring-segments illusion? Distribution of these two sets of scores with reference to each other is shown in Table II. Correlation between these two variables yields an epsilon coefficient of +0.79 with a standard error of 0.03 indicating a

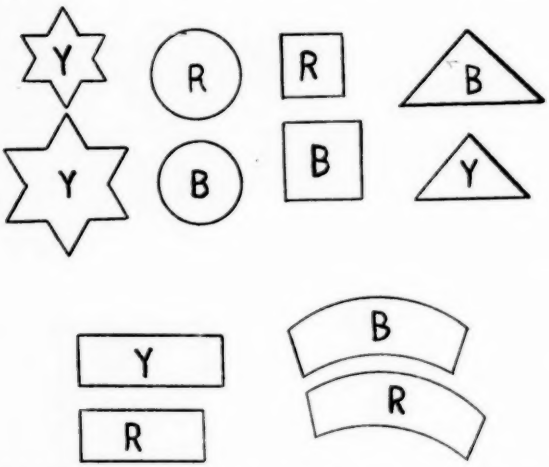


Fig. 1. Scaled representations of the six geometrical figures used in the experiment. Reading from left to right, each pair of figures is shown in the arrangement used for initial presentation. Y = yellow, R = red, B = blue. One-sixteenth inch equals one inch.

TABLE II

DISTRIBUTION OF SCORES ON CONCEPT TEST WITH REFERENCE TO RATING ON TEST FOR REACTION TO CONTRADICTORY RELATIONS

Rating of Reaction	Concept Score										Total
	3	4	5	6	7	8	9	10	11	12	
(+)	—	—	—	—	—	1	—	—	3	62	66
(?)	—	—	—	—	1	—	1	1	3	12	18
(−)	1	—	4	7	15	9	3	5	10	5	59
Total	1	0	4	7	16	10	4	6	16	79	143

very strong relation between the two variables. In addition, a phi coefficient was computed by dichotomizing concept test scores between 12 and 11 and combining (?) with (+) scores on the test for reaction to contradictory relations. The resulting dichotomies as entered in the fourfold table were as follows: positive on both tests, 74 children; negative on both, 54; positive on concept test with negative reaction, 5; negative on concept test with positive reaction, 10. The phi coefficient thus obtained, without correction for continuous distribution in one variable and broad categories, was +0.79, which thus gives a conservative underestimate of the strong relation between the two variables.

Proportion of children showing reaction to the illusion jumps from virtually none in the age group 2–3 to approximately 50 percent in the age group 3–4. The earliest age at which a reaction to contradictory relations appears, using the 50 percent criterion and the technique used in this study, may therefore be placed between three and four years depending, among other things, on occupational group. Thus, at 3–4, 53 percent of the high and 40 percent of the low group show positive reaction. At 5–6, 100 percent of the high and 76 percent of the low react to the illusion.

DISCUSSION OF FINDINGS

It has been shown that when the children demonstrated a functional use of the concepts employed, they also reacted to the contradictory relation posed by the ring-segments illusion in over 90 percent of the cases. This finding is also reciprocal. When it is considered that the unreliability of the two tests, the observers, and the children, serve to attenuate the true relation between the child's performance on the two tests, it is suggestive that the obtained coefficient of correlation may represent an attenuation of a true relation amounting to unity. Whatever the true relation may be, there is no comfort for the hypothesis that a relation of contradiction emerges as a later stage apart from concept formation or as an element later added to concepts. If the hypothesis is amended, however, to state that *solutions* to contradictory relations emerge at a later time dependent on the experience and native wit of the child, then there is nothing in the present study to refute it. Our findings agree quite well with Spearman's second noegenetic principle, the eduction of relations, insofar as that principle could be translated into behavior terms. On the basis of our findings we may venture to agree with Spearman that relations inhere in the very nature of the concept itself.

Piaget's placing the emergence of sensitivity to contradictory relations at seven or eight (3, 4) conflicts sharply with our finding a behavioral recognition between the ages of three and four. It may be that Piaget reached his conclusion more on the basis of the child's ability to reach a satisfactory solution of a problem than on a simple demonstration that he recognizes, however dimly, that an expected relation (belief, anticipatory schema) is disrupted, producing

thereby a sense of confusion, doubt or anti-expectedness. At the conclusion of testing, several of the children who had reacted positively to the illusion were asked, after several further reversals of the segments, which was the larger segment and how it was that the red segment was large and then small. The responses were generally characterized by suggestibility, self-contradiction and confusion. Some replies are given below, with age of child in parentheses.

I can't tell which because they're put together so. (4-1)
I just finished looking at the big one. (4-3)
It *looks* big. (4-6)
Now the blue one is the big one. (4-9)
When you do it that way, the blue is big. It makes the little one big. (4-10)
You changed it! How do you do it? (5-3)

So far as their verbal responses are concerned, it would seem that generally the incompatibility of big being little is hardly recognized; though one could infer with equal subjectivity that responses such as "It *looks* big" and "You changed it! How do you do it?" bespeak a fundamental recognition that big cannot be little, followed by a search for solution of the apparent discrepancy. In the *behavior* of the successful children, resolution of the conflict between present perception and previous judgment (solution of the problem) was not always 'logical,' and sheds an interesting light on the reality solution of the problem. Of the 79 children who scored perfectly on the concept test, 38 percent resolved the conflict by choosing the ring-segment which appeared larger at the time. 62 percent 'ignored' their present perception and chose the segment as large which they had previously judged to be larger.

Research suggestions.—The tendency of the two and three year old children to make the Big-and-Little problem a pick-up game rather than accepting the task set by the experimenter suggests that the role of what has variously been called set, *Aufgabe*, etc. (1), in the development of abstract concepts should be a fruitful field for investigation. It would suggest particularly the conditions for development of motivation in the direction of accepting various kinds of tasks as problems requiring solution and would possibly be especially relevant to 'ego development.'

It may be noted in Table II that a distinct bimodality in concept scores is restricted specifically to those children who showed no reaction to the contradictory relation. The two modes occur at chance (score of 7) and somewhat below a perfect score, while the distribution of scores for the successful children is unimodal with a perfect score of 12. This probably indicates that a well defined notion of relative size is essential to the perception or eduction of the contradictory relation. While the bases of error were not uniform for those scoring slightly below 12 points, a majority appeared to be troubled with the variable, order of request. Thus, for the circle and square, request both before and after reversal of the figures was for the same figure (big for circle, little for square), while order of request for all the other figures was first for the big and then for the little, or vice versa. The difficulty experienced by these children may be related to the factor of set mentioned above. This, together with the other bases of choice and the distractions and suggestions which hinder their perfect scoring is a problem requiring further investigation.

Most psychologists would no doubt agree with Terman that the child who shows a sensitivity to contradictory relations is to be regarded as

more intelligent than one who does not, other things being comparable. Since the test for reaction to contradictory relations appears to behave like an intelligence test item (sharp increase in proportion passing within a narrow age range, greater proportion of high occupational group passing, no sex difference), and hence shows some face, theoretical and practical, or predictive, validity, it is suggested that further investigation oriented towards its use in an intelligence test battery might be rewarding.

Summary and Conclusions

The problem was to determine whether the preschool child, upon demonstrating a functional use of the concepts *big* and *little*, would show at the same time a behavioral recognition of the contradiction when these two concepts are employed in a contradictory relation, i.e., whether this behavior emerged as an aspect of concept formation or as a later development.

Behavioral recognition was defined as behavioral reaction to the contradictory relation as an unexpected event. Ability of the child to use the concepts was determined by requiring size discrimination between pairs of geometrical figures. The contradictory situation was one pair of these figures, the ring-segments illusion which, upon reversal of position, results in the apparently larger segment becoming apparently smaller. Thus, a present perception is elicited which contradicts the child's immediately preceding perception and expressed judgment.

It was found that when the children demonstrated a functional use of the concepts employed, they also reacted to the contradictory relation posed by the ring-segments illusion in over 90 percent of the cases. This finding was also reciprocal, i.e., when they did not show a functional use of the concepts, they did not react to the illusion. On the basis of this evidence, it was suggested in agreement with Spearman that relations are essential aspects of concepts.

References

1. Dashiell, J. F. A neglected fourth dimension to psychological research. *Psychol. Rev.*, 1940, **47**, 289–305.
2. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.
3. Piaget, J. *The language and thought of the child.* New York: Harcourt, 1926.
4. Piaget, J. *Judgment and reasoning in the child.* New York: Harcourt, 1928.
5. Spearman, C. *The nature of intelligence and the principles of cognition.* London: Macmillan, 1927.
6. Terman, L. M. *The measurement of intelligence.* New York: Houghton Mifflin, 1916.
7. Terman, L. M., & Merrill, Maude. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

WORK AND REST AS VARIABLES IN CYCLICAL MOTOR LEARNING

BY GREGORY A. KIMBLE AND EDWARD A. BILODEAU

Brown University *

INTRODUCTION

Fortunately, a number of seemingly divergent theories developed to handle the typical superiority of distributed practice in motor learning have a fundamental point in common. They all postulate that during rest pauses there occurs a recovery from the effects of practice which are detrimental to performance. Key words in these various systems have been 'fatigue,' 'work decrement,' and 'inhibition.' In all of these formulations, performance is assumed to depend upon the strength of an excitatory tendency minus the strength of an inhibitory tendency. If we adopt the symbols of Hull (**5**), this postulation may be stated more formally as follows:

$$R = f_1({}_sE_R - I_R) \qquad (1)$$

In words, the strength of the response tendency is some subtractive function of the magnitudes of an excitatory potential and reactive inhibition.

Both ${}_sE_R$ and I_R are, in turn, postulated to depend upon independent, manipulatable variables. Our concern is only with the determiners of I_R. Several experiments (**2**, **4**, **6**, **7**, **8**) have shown that among the important variables determining the magnitude of I_R are the length of the work period (t_w) and the length of the time allowed for rest (t_r). The value of I_R at any moment, t, is determined by these two variables:

$$I_R t = f_2(t_w) - f_2(t_r) \qquad (2)$$

Since I_R increases as a function of the duration of practice and decreases with rest, this function is also to be considered subtractive.

In view of the fact that the separate effects of work and rest on learning are well understood, it might be supposed that we would know something about the joint effect of these two variables in the typical motor learning situation where work and rest are alternated in some orderly fashion. This, however, is not the case. The ordinary procedure in establishing the effect of two such variables on behavior, as outlined in standard sources on experimental methodology, would be to hold one variable constant and to vary the other systematically; then to hold the second constant and to vary the first. Furthermore, since magnitudes of both variables are dependent upon time, one would probably want to use a factorial design and sample equal values of each variable. This approach to the problem has, however, not been typical.

The main reason for this failure to use an appropriate experimental design is that too frequently investigation in the field of motor learning has been determined by available gadgets rather than the dictates of psychological theory and experimental convention. Very frequently the actual control of a pursuit rotor experiment, for example, is determined by the characteristics of a one-r.p.m. motor which automatically starts and stops the rotor at certain intervals. The basic consideration which determines the experimental procedure thus becomes the fact that the motor makes one complete revolution every minute. The experimenter may, for instance, in-

* The junior author is now at the State University of Iowa.

vestigate the course of learning under conditions where the *S* works for 30 sec. and rests for 30 sec., thus using up the available one min. unit. If he now decides to study the effect of increasing the length of the work period to 40 sec., he must reduce the length of the rest period to 20 sec. This subordination of experimental design to the physical characteristics of an instrument is a rather obvious methodological error. Some of the most extensive experiments on the distribution of practice, however, have committed it. One of the purposes of the present experiment was to study the effect of distribution of practice in a way which would avoid this difficulty.

So far in our summary of the decremental factor (I_R) in motor learning, we have identified the major variables, outlined their separate effects on performance and suggested a procedure for determining joint effects. Now let us ask some of the specific questions to which we hope to get answers from a simple experiment of this kind:

1. Which is the more important variable, work or rest, in determining the rate and amount of learning?
2. Is there an interaction between the two variables; i.e., does the effect of one variable depend upon the magnitude of the other?
3. Do the effects of the two variables cumulate with practice? If so, how?
4. How do the two variables cooperate in their effect on learning; does the effect of one variable merely add to the effect of the other, or is the joint effect more complicated?

The present experiment is, in a sense, a demonstration experiment set up to show how one might procede in getting the answers to these questions. On the basis of our results we are able to suggest answers to all of them. We are aware, however, that, because of the limited scope of the experiment, our answers are tentative. They may not apply equally well to the results of experiments in which much larger or much smaller values of the two experimental variables are sampled.

Experimental Technique

Apparatus

The apparatus employed was the Minnesota Rate of Manipulation Test. In the present experiment the test was used as a motor learning apparatus. This test consists of a large board with four rows of 15 circular holes. Into each hole a cylindrical block of wood may be fitted. *S*'s task was to overturn as many of these blocks as possible in a given period of time.

Subjects and Procedure

Ninety-six elementary psychology students were used as *S*s; 48 were men and 48 were women. Ages ranged from 17 to 30. All *S*s had previously had three practice trials on the apparatus in an experiment in the elementary psychology laboratory. Thus, the *S*s were familiar with the task but they were unfamiliar with the experimental design and procedure.

The experimental design involved two independent variables, length of working time and length of resting time. Two values of each were used: 10 sec. and 30 sec. work; and 10 sec. and 30 sec. rest. Thirty sec. was chosen as the longest trial because the standardization data on the test indicated that even the most rapid *S* would be unable to overturn all of the blocks in this length of time. The four experimental conditions fitted into a simple factorial design to allow for four experimental groups as shown in Table I.

TABLE I

The Design of the Experiment

Condition	Work Period	Rest Period	Reference Notation
1	10 sec.	10 sec.	10–10
2	10 sec.	30 sec.	10–30
3	30 sec.	10 sec.	30–10
4	30 sec.	30 sec.	30–30

Included in the last column of this table is the notation by which we will, for the sake of brevity, refer to each condition in our subsequent discussion.

Twenty-four *S*s were used in each of the four experimental conditions. In addition, the 24

were subdivided into 12 men and 12 women. Two *S*s, always of the same sex, were run simultaneously, each standing before his board and facing his partner. Thus, at any time either *S* could compare his performance with that of the other. The *S*s in the two 30-sec. practice conditions were given 10 trials; those in the two 10-sec. conditions were given 30 trials. The total practice time was, therefore, 300 sec. for all groups.

S's task was to overturn as many of the blocks as he was able between the measured interval of 'go' and 'stop.' Each *S* began with the block in the uppermost right hand corner, lifted it out with his left hand, transferred it to his right hand, overturned it, and replaced it into the same hole from which it had been lifted. *S* then proceeded as quickly as possible to the block to the left. At the end of the row, *S* began with the second row and the direction of movement was reversed.

The experimenter was seated so as to be able to survey both boards, time the work and rest intervals with a stop watch and record the number of blocks overturned on each trial. During the rest intervals, the *S*s stood before the boards with their arms at their sides and awaited the 'go' signal.

Results

The data obtained in the experiment were the number of blocks overturned on each trial by each of the 96 *S*s under the four experimental conditions. Two separate analyses of variance were performed to establish evidence for the effect of the work and rest variables at different points along the practice series:

1. The first analysis was made on the performance scores for the first 30 sec. of practice for the four experimental conditions. Scores for the two 10-sec. work conditions were obtained by combining scores for the first three 10-sec. learning trials. The scores for the two 30-sec. work conditions were the scores on trial one.

2. The second analysis of variance was performed on the performance scores of the last 30 sec. of work for all four experimental conditions. That is, on scores obtained after 270 sec. of practice. Again the data for the 10-sec. work condition are the sums of the scores for three 10-sec. work periods, but in this case, the last three trials. The data for the 30-sec. work conditions are the scores for the last trial.

The results of the analyses of variance are presented in Table II. The analysis on the data for the initial 30 sec. of learning is considered first. Here the *F*-ratio which tests for the significance of the performance difference resulting from a variation in the length of the work interval is significant beyond the one percent level of confidence. Thus, variation in the length of the learning trial has a pronounced effect upon performance almost as soon as practice has begun. This is not true for the rest variable. The *F*-ratio for rest is .14 and is not significant. The variance contributed by the interaction of work and rest is likewise found to be non-signifi-

TABLE II

Results of the Analysis of Variance Performed on Initial and Final Scores

Source of Variation	*d.f.*	Initial Scores		Final Scores	
		Mean Square	*F*	Mean Square	*F*
Between work periods	1	799.26	48.91*	810.9	58.68*
Between rest periods	1	2.34	.14**	231.3	16.74*
Interaction (work × rest)	1	1.27	.08**	3.7	.27**
Within groups (error)	92	16.34	—	13.82	—

* Significant beyond the one percent level of confidence
** Not statistically significant

cant. It would thus appear that performance differences obtained under our four conditions of distribution are solely the result of variation in the length of the practice trial.

The results of the analysis of variance performed on the data for the last 30 sec. of practice present a different picture. Estimates of the variance for *both* length of work period and length of rest period are significant beyond the one percent level of confidence. The interaction between these two variables, however, is not significant.

It is now apparent that with length of work period held constant, performance may be significantly bettered by lengthening the rest interval. With length of rest period held constant, increasing the length of work period significantly decreases work output. The non-significant interaction indicates that the trend in the effect of length of work period is similar from rest period to rest period, and the trend in the effect of rest period is similar from work period to work period.

In summary of the results of the analysis of variance, it may be said that in the early phases of learning, length of the work period is an active variable, clearly separating the performances of the 10-sec. and 30-sec. work conditions. But length of interpolated rest period does not function as an active variable in separating the work conditions until the later states of acquisition, where variation in either variable results in significant differences in performance.

The immediate effect of work and the increasing effect of rest are shown clearly in the learning curves for the four experimental conditions plotted in Fig. 1. The mean block scores for each condition have been plotted against seconds of practice at the ten

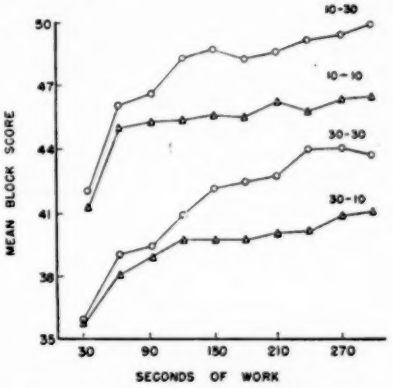


FIG. 1. Learning curves showing the performance of each of the four groups as a function of the number of seconds of practice. The scores for the two 30-sec. conditions were obtained directly. Scores for the two 10-sec. conditions are the mean scores for blocks of three trials.

sampled points along the practice continuum. It is evident that both 10-sec. work conditions were adequate samplings of a homogeneous population as were both 30-sec. work conditions, since their respective performances are quite similar during the first 30 sec. of work. The initial difference between the two work conditions is apparently the result of a 20-sec. difference in length of work period. The difference between the two 30-sec. work conditions is not marked until after 120 sec. of working time. From this point on the differences in mean score are clearly in favor of the condition with the longer rest period. In a similar manner the differences between the two 10-sec. work conditions appear more marked after 120 sec. of working time, and are again in favor of the condition with the longer interval of rest between trials.

Let us now turn to an analysis of the separate effects of rest and work upon performance. In Fig. 2, the curve with the triangles represents the

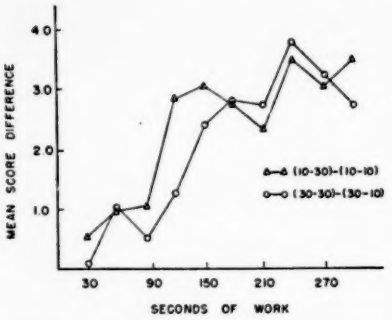


FIG. 2. Performance differences as a function of a 20-sec. difference in the length of the rest interval with the length of the work period constant at 10 sec. and 30 sec. Plotted points are successive differences in the learning curves shown in Fig. 1.

mean scores of the 10–10 condition subtracted from the mean scores trial by trial for the 10–30 condition. This has been done in order to present the successive performance differences which result from increasing the length of the rest period by 20 sec. while holding the length of the work period constant at 10 sec. Using an identical procedure, the curve with the open circles represents successive performance differences between the 30–10 and 30–30 groups. As in the case of the curve mentioned above, this curve shows the effect of increasing the rest interval by 20 sec. But in this case, the length of the work period is 30 sec. for the two groups. Of particular interest is the fact that the two curves cross each other at three points and appear to run a similar course. At least for short work periods, it appears that increasing the length of the rest period results in similar performance gains throughout the course of learning regardless of the absolute length of the work periods involved. This fact has been anticipated by the statistical finding that there was no interaction between our two variables either at the beginning or at the end of the learning series.

The analysis of the preceding paragraph is illustrative of the further analysis that we shall now attempt to make of the separate and joint effects of work and rest upon motor learning. The technique which will be used involves a subtraction of the mean scores in one condition from those in a second condition at each of the 10 sampled points in learning shown in Fig. 1. Such an analysis is proper because of the fact that the experiment was designed to employ equal periods of work and rest. The results of this procedure are presented in Fig. 3. Observe first of all that these curves are all obtained by subtracting the mean scores of other groups from those of the 10–30 condition. Taking this condition as a reference point, what these curves show us is the trial-by-trial effect of changing one or both of the experimental variables by 20 sec.

The bottom curve shows the effect of decreasing the length of the rest

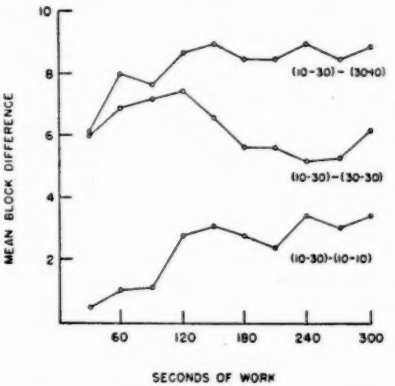


FIG. 3. Performance differences resulting from: 1. shortening the rest period by 20 sec. (bottom curve); 2. lengthening the work period by 20 sec. (middle curve); 3. shortening the rest period and lengthening the work period both by 20 sec. (upper curve).

period by 20 sec. The result is an increasing score difference in favor of the condition allowing the longer rest. The intermediate curve shows the effect of increasing the length of the work period by 20 sec. The most striking feature of this curve is the immediacy of the score difference produced by changing the length of the work period. Beyond this point, the curve appears to have a sinusoidal character. However, in terms of score differences, the fluctuations are small. For theoretical reasons, to be considered later in the paper, we are inclined to view this curve as a simple straight line of zero slope.

The top curve shows the effect of shortening the rest period and lengthening the work period simultaneously. The result is a curve which resembles the bottom curve in shape but which is displaced above it by an amount approximately equal to the score difference contributed by variation in the work variable (middle curve). This immediately suggests that the joint effect of work and rest is a simple, additive function of the two separate effects. Again the experimental design makes a test of this hypothesis possible.

Note that, in Fig. 3, the bottom curve shows the effect of shortening the rest period by 20 sec. and the middle curve shows the effect of lengthening the work period by 20 sec. The top curve represents the effect of changing both work and rest by 20 sec. In other words, the differences represented in the top curve are the equivalent of those represented in the two lower curves. If it should happen to be true that the effects of work and rest are additive, the summation of the values which go to make up the points on the two lower curves trial by trial should produce a curve which coincides with the top curve.

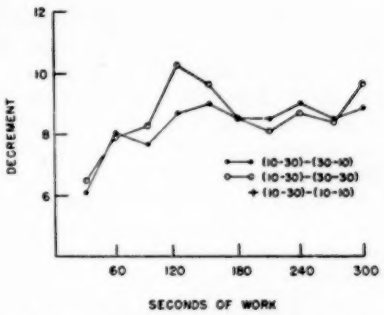


FIG. 4. Graphical test of the hypothesis that the separate effects of work and rest are additive in producing their joint effect. For the way in which the points were determined, see text.

The result of performing this summation is shown in Fig. 4. Here the curve with the solid circles is the top curve of Fig. 3 and the curve with the open circles represents the values obtained by the addition of the value in the two lower curves. For nine of the ten points the two curves are nearly identical.

DISCUSSION

Our discussion of the results of the experiment will consist of an attempt to answer the four questions the investigation was designed to answer.

1. *Which is the more important variable, work or rest, in determining the rate and amount of learning?* On the basis of the experimental results obtained, the answer seems to be the amount of work. In Fig. 1 we saw that the two conditions with the 30-sec. work period were the lowest in terms of performance. In Fig. 3, we saw that the score differences resulting from varying the work period were greater than those resulting from a variation in the amount of rest. In terms of the development and decay of I_R, this result strongly suggests that the process of development is much more rapid than that of decay.

2. *Is there an interaction between the two variables;* i.e., *does the effect of one variable depend upon the value of the other?* For the limited range of amounts of work and rest sampled in this experiment, the answer to this question seems to be 'no.' Both our statistical and graphical analyses point in this direction. Even on the last 30 sec. of learning where performance was apparently at a maximum, no evidence for such a relationship could be found. If this finding is borne out by the results of experiments covering a wider range of work and rest intervals, the analysis of motor learning turns out to be somewhat simpler than we would have predicted, because the effect of each variable will have been shown to be independent of the other.

3. *Do the effects of the two variables cumulate with practice? If so, how?* Here our tentative answer depends upon the particular variable in question. For the rest variable, the effect seems to increase with practice. The precise shape of the curve is somewhat indefinite. Reference to Fig. 2 or to Fig. 3, where such functions are plotted, will show that either a negatively accelerated or an S-shaped curve would fit the data fairly well.

For the work variable, the answer is still more uncertain. Taken at their face value, our results would indicate a wave-like function. Our preference, however, is for a straight line function of zero slope which would indicate that the effect of increasing the length of the practice trial is to produce an immediate and constant reduction in the rate of performance. Our preference for such a function is based chiefly upon existing evidence that under conditions of continuous practice, a constant amount of I_R develops early in learning and is maintained at a fairly constant level throughout learning. The logic behind this point of view is to be found in papers by Ammons (1) and Kimble (7). Both of these investigators have presented evidence which partially confirms the hypothesis (2, 7). At the same time, however, the attainment of a constant level of I_R in only 30 sec. of practice would not be expected on the basis of the evidence available. The results of Ammons and of Kimble indicate that the attainment of this level may require as much as seven or eight min. of continuous practice and that two or three min. is probably the absolute minimum in the continuous practice situtation. What is clearly needed is further experimentation to establish the effect of varying the length of the work period more definitely. If on the basis of such research, our hypothesis that lengthening the work period produces an immediate and constant work decrement is confirmed, it will probably mean that cyclical motor learning requires a different type of analysis from that which has been applied to the case in which continuous work and a single rest are considered.[1]

4. *How do the two variables cooperate in their effect on learning?* Perhaps the most significant result of the experiment is to be found in our answer to this question. On the basis of the rather close agreement of the two curves presented in Fig. 4, it would appear that the independent effects of work and rest summate in a simple fashion to produce their joint effect.

[1] In a paper which is now in preparation, an analysis is made of some of the results reported by Kientzle (6) and some others obtained by Kimble. The choice of work and rest intervals happened to be such that an analysis following the line of that presented in the present paper can be performed. The results of such an analysis show quite clearly that the lengthening of the work interval produced an immediate and constant decrement in performance.

The significance of this finding is that it serves to simplify our analysis of the process of motor learning. Our results have shown rather conclusively that a knowledge of the separate effects of the two variables enables us to predict the effect on behavior of varying them together.

Summary

1. Four conditions of cyclical alternation of distribution of practice were each administered to 24 Ss in a rate of manipulation task in an investigation of performance as a function of length of work period, and as a function of length of interpolated rest. The four conditions employed were the following: 10 sec. work, 10 sec. rest; 10 sec. work, 30 sec. rest; 30 sec. work, 30 sec. rest; 30 sec. work, 10 sec. rest.

2. The rank order of the final degree of learning is the following: a. 10 sec. work, 30 sec. rest, b. 10 sec. work, 10 sec. rest, c. 30 sec. work, 30 sec. rest, d. 30 sec. work, 10 sec. rest.

3. It follows from this that the largest performance difference is a function of the length of the work period.

4. Performance differences as a function of the work variable appear to be constant. The effect of varying the length of the work period appears early in learning and is maintained at a nearly constant level throughout practice.

5. Performance differences as a function of the length of the interpolated rest increase with practice.

6. The joint effect produced by the concomitant lengthening of the work period and shortening of the rest period is a simple summation of the two separate effects.

(Manuscript received February 25, 1948)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. Bilodeau, E. A. Work and rest as variables in motor learning. Unpublished Sc.M. thesis. (On file, Brown Univ. Library)
4. Ellson, D. G. Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton Century, 1943.
6. Kientzle, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.
7. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
8. Kimble, G. A. & Horenstein, B. Reminiscence as a function of the length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.

THE GUESSING-SEQUENCE HYPOTHESIS, THE 'SPREAD OF EFFECT,' AND NUMBER-GUESSING HABITS

BY WILLIAM O. JENKINS

Harvard University

AND

LETA M. CUNNINGHAM

Indiana University

Purpose of the Studies

There were two major purposes of the present studies: (1) to perform a critical test of the guessing-sequence hypothesis (1) as contrasted with the Thorndike explanation of the 'spread of effect' (4, 5); and (2) to determine the extent to which sequences of numbers guessed freely show a stereotyped pattern for all *S*s.

In brief, the 'spread of effect' is typified by a bi-directional gradient of repetitions of incorrect responses neighboring correct or rewarded responses where *S*s are instructed to guess a number to each of a series of individually presented word-stimuli. As originally described by Thorndike (4, 5), the frequency of repetition of the non-rewarded responses decreases with increasing remoteness of the errors from the rewarded response in both a preceding and following direction. Thorndike interpreted the gradient as the 'spread' of the action of reward from the reinforced connection to adjacent stimulus-response bonds, and presented the findings as proof of the automatic effects of reward.

A considerably more parsimonious explanation has been proposed by Jenkins and Sheffield (1) whose evidence strongly suggested that *S*s' habits of guessing numbers were the main variable accounting for the 'spread-of-effect' gradient following reward. The guessing-sequence hypothesis explains the fact that a declining after-gradient is not found if the rewarded response is not repeated as is shown in the findings of these investigators. According to this hypothesis, a given *S* tends to guess numbers in a sequence that is predictable. For example, he may consistently guess '4' after having said '5.' If the preceding number is repeated to the same word the next time through the list for any reason, the succeeding number is likely to be repeated because of the predisposition to guess '4' after '5.' Since the correct responses tend to be learned and repeated, the same sequence of numbers is likely to be given following a correct number. It is also likely that the strength of the guessing tendency will decrease with increasing remoteness from the 'anchored' (rewarded) response because number habits are not perfectly predictable. Two conditions are necessary for this hypothesis to account for the gradient following reward: (1) an apriori tendency for *S*s to guess numbers in a sequence, and (2) that the rewarded response be repeated.

Experiment I. Test of the Conflicting Hypotheses

According to Thorndike's hypothesis concerning the automatic effect of reinforcement, *E*'s saying 'right' to a particular number response should spread to 'stamp in' neighboring word-number bonds. It follows that increased repetition of these incorrect numbers in response to the same words should occur *regardless of the position of the word in the list* on succeeding presentations. The gradient, then, should be found according to

the original remoteness of the S-R 'bond' from reinforcement. Contrariwise, from the guessing-sequence hypothesis the prediction may be made that the gradient of repetition will follow the correct, repeated response *regardless of the stimulus words* where a different order of presentation of the list is employed on successive trials. That is, if the rewarded response is repeated, repetition of the number responses following reward can be expected without reference to the stimulus-words involved.

There is no reason to expect from Thorndike's theory that changing the word order would destroy the gradient. From the guessing-sequence hypothesis, however, no gradient in the responses to incorrect words after the rewarded word is to be expected where the presence of reward has occurred on the first trial. But a gradient for position would be expected in the incorrect responses after the rewarded number where the reward has occurred on both the first and second trials.

A test similar to the present one has been performed by Zirkle (8) in a different context. This investigator, however, and many of the previous students of the 'spread of effect,' failed to demonstrate the existence of a gradient statistically, as pointed out by Tilton (6). In addition, the level of repetition of the rewarded responses was atypically low in Zirkle's results, which are treated in the discussion section of this paper.

Experimental Design and Procedure

The experimental test of the conflicting predictions from the Thorndike and guessing-sequence interpretations was set up as follows. The typical situation for measuring 'spread of effect' was employed involving the guessing of a number to each of a series of individually presented words. Since the data to check both predictions were available from a single testing of each *S*, the procedure was the same for both conditions. The words consisted of 21 three-letter nouns each printed on a 3×5 in. flash card. Each *S* was provided with a record sheet on which to record responses, and a cover card to prevent seeing previously recorded numbers. Presentation of the flash cards individually to the group of *S*s was accomplished by *E* at the rate of one every three sec.

One deviation from the Thorndike procedure was to employ a fixed point of reward. That is, at three previously selected places in the list the same words appeared on all trials and *E* said 'repeat' after each of the three words. This group procedure was taken over from Smith (2). The *S*s were instructed to learn and to repeat their responses to these three stimuli on succeeding trials. This procedure of a fixed reward avoids the generation of a spurious gradient before reward which Tilton (7) has pointed out results in the typical Thorndike set-up.

Another deviation from the standard method which was basic to the design of the experiment consisted in systematically revising the order of the words in the list on each succeeding presentation. The list of 21 words was divided into three groups of seven each. The middle item of the seven was held constant and the response to it was 'rewarded' in every case by the saying of 'repeat,' with the responses to the three words preceding and following reward going unreinforced. The rules for reordering the list of words were: (1) No two words occurred together on successive presentations of the list. (2) No incorrect response occurred in the same group of seven on the succeeding trial. (3) Incorrect responses before reward on one trial occurred after reward on the succeeding trial, while those after reward on one trial were placed before it on the next presentation. (4) The positions of the three words to which *E* said 'repeat' were held constant on all five trials.

The instructions to the *S*s as a group were as follows:

> This is an experiment in learning. You will be shown a series of 21 words, one at a time, on these cards. The series of words will be presented five times. To each word you are to respond by writing any number from one to ten on a separate line of your record sheet. The numbers one and ten are to be used. Do not use any system of writing numbers, but respond as randomly as possible. After you have written a number down, cover it with the card and wait for the presentation of the next word. Do not look back at the numbers you have just written. Write only one number for each word.

After you have written numbers to certain words I will say 'repeat.' Try to learn and remember that the number you have put down goes with the word and try to respond with the same number the next time you see that word.

When the question was asked, the *Ss* were told that the words would not necessarily be presented in the same order each time.

A total of 50 *Ss*, members of two undergraduate classes in psychology, were tested. The records of 10 of these were discarded because they gave either more or less responses than the 21 required and it was not possible to find which responses were intended for which words. The records, then, of 40 *Ss* were employed in the treatment of the data.

The statistical measure of significance used was the *t*-test for correlated data. Since the same *Ss* had been tested under both conditions, and since, by the nature of the procedure used, the total number of responses at all points was equal, it was possible to compute the number of repetitions for each *S* at each unit position for each experimental condition separately and work with difference scores thus taking the correlation element into account. Percent repetition could be calculated by dividing repeated responses by a constant, the total number of responses.

The analysis procedure for the gradient predicted by the guessing-sequence hypothesis consisted of recording the number of repeated responses in positions three before reward, two before, one before, one after, and so forth, without regard to the stimulus words involved. The number repeated at the various positions for all five trials combined was converted to a percent by dividing it by the total number of responses. Since a basic condition for the operation of the guessing-sequence variable is that incorrect responses after reward be 'anchored' by the repetition of correct ones, the main data were the frequency of repetition of these errors around repeated, rewarded responses.

The analysis of the data according to Thorndike's position consisted of determining the position of a stimulus word and its number response on a succeeding trial and tabulating the response as a repetition or non-repetition in the position around reward held by that word on the preceding trial. Thus, a word that occurred in the first position before reward on the first trial might be in the third position after reward (in another group) on the succeeding trial. The response to this word on the second trial would be recorded as a repetition or non-repetition by referring to the number given to the word on the preceding trial. This particular response would be tallied in position one before reward, that is, according to its position on the first of the pair of trials.

In this connection the critical point statistically is the demonstration of a gradient effect. To show a gradient for either experimental condition, frequency of repetition at unit position one after reward must be reliably higher than that at position two after reward. Likewise, percent of repetition at position one before reward must be significantly higher than that at position two before reward.

Results

Fig. 1 and Table I present the data most relevant to the guessing-sequence hypothesis, namely, the percent of repetition for the three unit positions preceding and following repeated, rewarded responses. The results are given for both the Thorndike prediction and the expected outcome from the guessing-sequence hypothesis.

A relatively pronounced gradient following reward in the guessing-sequence data can be seen from the

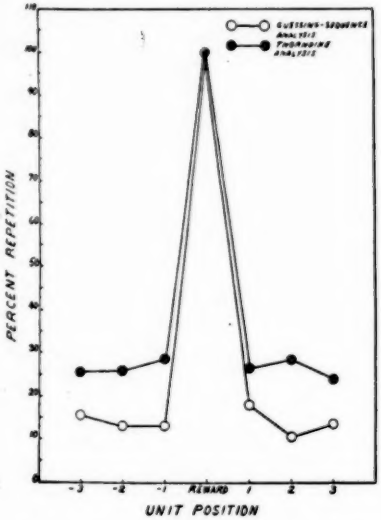


Fig. 1. 'Spread-of-effect' gradients around repeated, rewarded responses for the Thorndike and guessing-sequence analyses. Each point is based on 376 responses of 40 *Ss*.

TABLE I

'SPREAD OF EFFECT' AROUND REPEATED, REWARDED RESPONSES

A comparison of the percentage of repetition of responses in the designated positions with respect to repeated, rewarded responses for the Thorndike and guessing-sequence analyses. The percents are based on 376 responses of 40 *Ss* at each point.

	3 Before	2 Before	1 Before	Reward	1 After	2 After	3 After
Guessing-sequence analysis	15.4	13.0	13.0	100.0	18.1	10.6	13.6
Thorndike analysis	25.5	26.1	28.9	100.0	26.6	28.9	24.2

figure. Also apparent are slight gradient effects preceding and following reward in the Thorndike results. Statistical treatment of the latter, based on an *N* of 40, yielded no *t*-values exceeding 1.2 for the differences in number of repetitions at the various points. None of these values attains the 10 percent level of significance.

In the guessing-sequence data, none of the differences between points before reward reaches the 10 percent level of confidence. The *t*-value for the difference between unit positions one and two following reward is 2.5 with a *P*-value of slightly less than .01.

Combining either points one and two or two and three both before and after reward and contrasting the data with the remaining results does not change the conclusions to be drawn from the statistical analysis.

From this treatment, the existence of a gradient after reward in the guessing-sequence data is supported. No other gradient effects are apparent. It might be noted that the level of repetition in the Thorndike data is higher at better than the one percent level of significance at all unit positions except one after reward where the *P*-value is .10.

In Table II are shown the results for the other part of the data, namely, the frequency of repetition of incorrect responses around non-repeated, rewarded responses.

None of the differences among gradient points within either set of data is significant at better than the five percent level of confidence. The level of repetition before reward in the Thorndike data is higher at the five percent level of significance than that in the corresponding guessing-sequence findings. The data for the two conditions are comparable after reward. Statistical support is lacking for the existence of any gradient effects in Table II.

TABLE II

'SPREAD OF EFFECT' AROUND NON-REPEATED, REWARDED RESPONSES

A comparison of the percentage of repetition of responses in the designated positions with respect to non-repeated, rewarded responses for the Thorndike and guessing-sequence analyses. The percents are based on 104 responses of 40 *Ss* at each point.

	3 Before	2 Before	1 Before	Reward	1 After	2 After	3 After
Guessing-sequence analysis	10.6	8.6	14.0	0.0	11.5	17.3	8.6
Thorndike analysis	23.1	15.4	18.2	0.0	13.4	14.4	14.4

TABLE III

'Spread of Effect' Around All Rewarded Responses

A comparison of the percentage of repetition of responses in the designated positions with respect to all rewarded responses, repeated and non-repeated, for the Thorndik and guessing-sequence analyses. The percents are based on 480 responses of 40 *Ss* at each point.

	3 Before	2 Before	1 Before	Reward	1 After	2 After	3 After
Guessing-sequence analysis	14.4	12.1	13.3	78.3	16.7	12.1	12.3
Thorndike analysis	25.0	23.8	26.7	78.3	23.8	25.8	22.1

Another comparison to be considered is that between the level of repetition of non-rewarded responses around repeated as contrasted with non-repeated, rewarded responses in the Thorndike data. From an inspection of Tables I and II, it can be seen that the percentages of repetition are considerably higher around repeated, rewarded responses. The over-all difference is significant beyond the one percent level of confidence. A further breakdown of the data was performed comparing the results for the first pair of trials with those for trials two through five. While the findings did not reach an acceptable level of significance, it was found that the differences were almost doubled in the later part of the trials.

In Table III are summarized the data for repeated and non-repeated rewarded responses combined for both experimental conditions. These data are comparable to the findings reported by previous investigators (4, 5, 8).

Within the two sets of data in this table there is only one difference that approaches statistical significance. That is between unit positions one and two after reward in the guessing-sequence findings. The t for this difference is 1.8 with a P-value of .04. There is no statistical evidence for gradient effects preceding reward in either set of data, and none for a post-reward gradient in the Thorndike results. The frequency of repetition in the Thorndike findings is appreciably higher than in the guessing-sequence results in line with the data reported in Tables I and II. The differences exceed the five percent level of significance in every case.

Discussion

This study tested the existence of a gradient predicted by Thorndike's theoretical position as compared with the gradient after reward expected from the number-guessing habits of *Ss*. An examination of the data in the previous section suggests the following points.

When the analysis was made with regard to position as follows from the guessing-sequence hypothesis, a gradient was demonstrated following reward which met a rigorous standard of statistical significance. Thus, the prediction was borne out: the guessing habits of *Ss* operate to produce a gradient after reward even though different stimulus words are involved on succeeding trials. No statistically acceptable gradient effect following reward was apparent in the Thorndike data. Nor was any reliable gradient found before reward in either set of results.

It appears that, at least for the gradient following reward in the 'spread-

of-effect' situation, the gradient is generated by *S*s' number-guessing habits anchored by the repetition of rewarded responses. The evidence suggests that an after gradient does not follow from the automatic action of reward in 'stamping in' S-R connections.

It is relevant to note that Zirkle's (8) results are in line with the present findings. He followed the procedure of mixing the order of presentation of stimulus items on successive trials and measured the effect with regard to both position and original stimulus on the following trial as in the present study. The trends of Zirkle's results are in the same directions as the present data exhibiting no gradients around reward for the Thorndike situation, a slight before gradient and a fairly pronounced after gradient for the analysis by position. It should be noted that Zirkle's treatment involved all rewarded responses whereas the guessing-sequence hypothesis assumes the *repetition* of a rewarded response as a basic condition for the action of guessing habits. Unfortunately this investigator failed to demonstrate the statistical existence of gradient effects. The present writers checked his findings and were unable to discover evidence for a gradient in any case attaining the five percent level of significance.

In the original study of the guessing-sequence hypothesis (1), the mechanism for generating a gradient before reward was suggested. When the repeated rewarded responses are selected out from the non-repeated ones, they probably include a number that are repeated not solely through the action of learning in the experimental situation, but rather because they are 'anchored' by the repetition of the response to the stimulus immediately preceding the reward point. For example, if the word *egg* occurs in unit position one before reward and there is a prior association with the number *2* so that a supra-chance tendency to say *2* after *egg* exists; and if the word *dog* is in the reward position, the response *3* may be given to *dog* and *repeated* not because of learning, but because *S* tends to say *3* after *2* and tends to repeat *2* to *egg* just before reward.

From this position it follows that, with a fixed point of reward and a constant order of presentation of the list items, there should be at least a slight gradient preceding reward in the case where repeated, rewarded responses are partialled out from non-repeated, rewarded responses. Such a prediction is borne out in Taylor's (3) data. It is of interest to note that no gradient preceding reward appears in the present analysis from the guessing-sequence approach. In the treatment from Thorndike's position, however, a slight and statistically insignificant gradient before repeated, rewarded responses is suggested by the data.

By comparable reasoning, one would expect an *inverse* gradient both preceding and following reward in the case of non-repeated, rewarded responses. This prediction is not supported by Taylor's (3) findings, which suggest relatively constant frequency of repetition for four unit positions preceding and following non-repeated, rewarded responses. The present data are not legitimately treated in this fashion since the order of presentation of the items was systematically varied from trial to trial. A basic assumption of the present line of reasoning is the necessity for the same response to be given to the same stimulus word in unit position one preceding reward on successive trials.

One other point might be made here. It will be recalled that the level of repetition in the Thorndike data was appreciably, and in most cases significantly, higher than that in the guessing-sequence results. This finding does not change any of the interpretations since the predictions were made with regard to the appearance or non-appearance of gradient effects and not the absolute level of repetition which would result. One obvious possibility is that the words employed (such as *ham*, *wig*, *foe*, *lot*, and *ice*) provided simple bases for forming easily learned associations or for the action of previously learned predispositions to respond with certain numbers to specific words. Thus, a number might be repeated to a given word with a high frequency regardless of the position of the word in the list on successive presentations.

Another finding to be noted is the large difference between repetition around repeated as opposed to non-repeated, rewarded responses in the Thorndike data. One possibility is that distraction in the form of the constantly changing order of presentation of the stimuli was operating. Another distraction that fits in with previous evidence (1) is the occurrence of repeatedly correct responses. The only constant feature of the test situation was the rewarded words. Reproduction of a correctly repeated, rewarded response and rehearsal of it may have disrupted the 'unlearning' of neighboring, incorrect responses acting differentially to result in the higher level of repetition around repeated as contrasted with non-repeated, rewarded responses.

Experiment II. Group Number-Guessing Sequences

One question to ask in connection with the action of the guessing-sequence hypothesis as an explanation of the 'spread-of-effect' phenomenon is whether *S*s as a group exhibit tendencies to guess numbers in a systematic fashion. It will be recalled that the guessing-sequence hypothesis does not require a group tendency in this regard, but only postulates that given any individual predisposition to say one number after another and given the anchoring of the first number in any way (such as by *E*'s saying 'right'), the gradient after reward will be generated. If, however, a group tendency can be shown, it is strong presumptive evidence for the case in point and will save considerable effort in the treatment of data.

The purpose of the present study was to determine group tendencies to guess numbers in a systematic fashion in a situation where *S*s were motivated by the instructions to respond randomly.

Experimental Design and Procedure

Two samples of undergraduate *S*s were tested in this experiment. One sample included 43 of the 50 cases tested in the first experiment. The second group was an independent sample of 34 *S*s. Data for a total of 77 *S*s are reported in the results.

Each *S* was presented with several strips of lined paper and a cover card and instructed to write down numbers in a random order as rapidly as possible in accordance with the following instructions.

> This is an experiment in speed of reaction. You have been given sheets of lined paper on which to write. When I give the signal you are to write the numbers from 1 to 10 in a random order, one number to each line. The numbers 1 and 10 are to be used, but not 0 or 11. Do not use any system of guessing such as 1, 2, 3, 4, 5, 6, and so forth. Guess the numbers as randomly as possible. See how many numbers you can write before the signal to stop is given. You will have about two minutes in which to work.

For each *S* a minimum of 80 numbers was obtained within the time limit set. A few *S*s

were able to record as many as 125 numbers in the two minutes.

The data were combined for all *Ss* in a ten-fold scatter plot. From this table percentages of frequency of occurrence of each number after every other were calculated by dividing the obtained frequency in a given cell by the total number of times the given number occurred. Following this, the data for all numbers were combined by dividing succeeding numbers occurring a given number of units from a preceding response by the total number of responses possible for that position. This procedure is essentially a contingency technique which summarizes the frequency of occurrence of numbers differing by a specified amount from any number between one and ten. For example, a number five less than another number could potentially occur for the case where the preceding number was 10, 9, 8, 7, or 6. Likewise, a number five larger than a preceding number could be given when the latter was 1, 2, 3, 4, or 5. For various positions the base *N* was the total cumulative number of preceding responses for all numbers involved. By combining the frequency of occurrence of each of the possible deviations, percentages could be computed. Thus, in the data presented below, the figure given for the position −9 from a given number must be all cases of 1 following 10. Similarly, all numbers under the heading −8 must be either a response of 2 after 10, or 1 after 9. The cells and categories were, of course, mutually exclusive. That is to say, the data entering into the computations of −9 were not involved in calculating the data for position −8.

The *t*-test for correlated data was employed for measuring the significance of the obtained differences in frequency of occurrence.

Results

The basic data for the 77 *Ss* are summarized in Table IV which presents the frequency in percent of the various number responses. These percents were calculated by summing the number of responses across rows and dividing the number in each cell in a given row by the total. The rows, then, add up to one hundred percent. The total number of responses on which the percents are based is given at the bottom of the table.

Several items are immediately noteworthy from Table IV. For one thing, the frequency of occurrence is greatest for the numbers immediately preceding and immediately following the number which has just been given, except, of course, in the case of 1 and 10 for which responses one removed can occur in only one direction. The modal value for the other eight numbers in Table IV can be seen to be the number immediately preceding. A

TABLE IV

Frequency of Number-Guessing Responses

The frequency in percent with which 77 *Ss* guessed a total of 7492 numbers from 1 to 10 at 'random.' The percents have been rounded to the nearest whole number.

Preceding Number	Following Number									
	1	2	3	4	5	6	7	8	9	10
10	24	9	8	6	4	5	8	13	19	3
9	6	5	7	5	4	7	12	30	1	23
8	5	7	6	7	7	13	23	1	18	13
7	4	4	6	10	9	25	1	19	18	5
6	4	6	10	13	22	1	15	18	9	3
5	5	9	15	23	1	15	14	9	5	4
4	9	12	32	1	13	11	9	5	5	2
3	14	28	1	14	13	7	7	7	7	3
2	37	1	15	12	8	7	6	5	7	2
1	1	14	18	13	11	7	12	9	8	8
Number of responses	729	714	861	781	708	789	845	839	735	487

second mode appears most typically at the value one following a given number. For all responses about 27 percent of succeeding numbers were one less than the number just given. Excluding 1 and 10, 43 percent of responses were within plus and minus one unit of a given number. Omitting 1, 2, 9, and 10, over 70 percent of the succeeding responses given fell within the range of plus or minus two units of the preceding number-response. Another item which may be seen in Table IV is the decided tendency for a decreasing frequency as the following number diverges from the preceding one.

The data of Table IV indicate, generally, a non-random order of guessing by the *Ss* involved, with a pronounced tendency to give numbers close to the number just given, but rarely a response corresponding to the preceding number.

The general predispositions just noted are brought out more strikingly in Table V and Fig. 2. As noted previously the findings reported in Fig. 2 and Table V were obtained by dividing the frequency of a given number-response by the maximum possible for a particular position. Thus, for position 9 the 58 responses of the 735 possible indicate that the response '10' was given 58 times after the response '1' had been made, or eight percent of the possible occurrences.

The orderliness of the data in Fig. 2 and Table V is apparent. The data for the two samples separately differed at no point by more than four percent except for position −9 where the difference was seven percent. None of these differences reached the 10 percent level of significance. The high frequency at position −9 is attributable to the pronounced tendency for 1 to be given after 10. Such responses might be considered as corresponding to the giving of 11 after 10. A lesser tendency to say 10 after 1 is apparent at position 9. It is unlikely that these trends would have appeared if only the middle values of a wider range of numbers had been employed.

The *t*-value for the differences between the related percents at each of the following points was significant beyond the five percent level of confidence: −1 and all other points except −9; 1 and all positive points except 2; −2 and −5; 2 and 5; and −9 and all other points except −1, 1,

TABLE V

FREQUENCY OF NUMBERS GUESSED IN DEVIATION UNITS

The frequency in percent with which 77 *Ss* guessed a total of 7492 numbers 'randomly' from 1 to 10, expressed in deviation units from the preceding responses. 'Number of responses' refers to the frequency of numbers differing by the specified amount; 'Base *N*' is the number of possible times such responses could have occurred.

Deviation	Number of Responses	Base *N*	Percent	Deviation	Number of Responses	Base *N*	Percent
−9	115	487	23.5	1	1126	6982	16.1
−8	91	1217	7.6	2	924	6252	14.8
−7	112	2054	5.4	3	450	5415	8.3
−6	166	2899	5.7	4	297	4570	6.5
−5	174	3688	4.6	5	223	3781	5.9
−4	242	4396	5.5	6	204	3073	6.7
−3	442	5177	8.5	7	138	2292	6.0
−2	758	6038	12.5	8	77	1431	5.4
−1	1805	6740	26.9	9	58	735	8.0
0	89	7492	1.1				

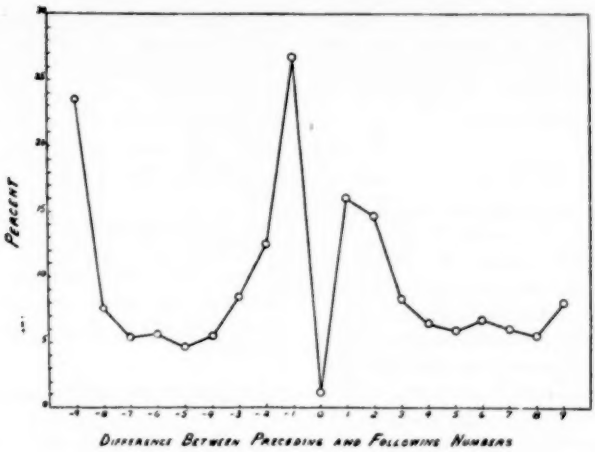


FIG. 2. Frequency in percent with which 77 *S*s guessed a total of 7492 numbers expressed in deviation units between the preceding and following numbers

and 2. These results indicate significant drops from the three major peaks of Fig. 2 to the neighboring depressions.

DISCUSSION

The present evidence suggests a strong group tendency for *S*s to guess numbers in a predictable fashion even though they have been instructed to respond in a random order. The data further suggest the direction of the tendency: succeeding number responses are close in unit value to the number previously given. The most pronounced trend is to give a number one unit less than the response just previously made. It is rare that the same number follows itself. Nearly half of all the responses fall within plus or minus one of the previous number, and over two-thirds are included within the limits of plus or minus two units of the preceding response.

It will be recalled that the guessing-sequence hypothesis makes no assumptions as to group tendencies to respond with numbers in a sequence. It only postulates a tendency for individuals to have a predisposition (from any source) to say one number after another. One *S* might say *7* after *1* while another might give *8* after *3*. The present results suggest that such connections are relatively infrequent. It should be noted that the atypical associations (say 2–9) could still operate to produce the after gradient of the 'spread of effect' once the preceding number was anchored by *E*'s saying 'right' and repeated.

It is not unreasonable to expect that number-guessing sequences exhibit the effects of previous training. For example, in their early schooling most individuals learn the number series from one on in order in the operation of addition. Somewhat later they learn the multiplication tables. Strong predispositions of several different kinds to respond with numbers in a sequence may well be built up and exhibited where *S* is not given special training in the meaning of the

word 'random.' It seems unlikely that a group of mathematical statisticians would show behavior patterns in guessing numbers comparable to those of the present *Ss*.

Summary

1. The present studies were undertaken to test Thorndike's theory of the automatic action of reward as compared with the guessing-sequence hypothesis as alternative explanations of the 'spread of effect'; and to determine group tendencies in guessing numbers sequentially.

2. In the first experiment, the order of presentation of the stimulus words in the typical word-number situation was systematically varied. The guessing-sequence hypothesis predicts a gradient following repeated, rewarded responses regardless of the stimulus on successive trials. It follows from Thorndike's position that the gradient should appear around the word-number associations according to their proximity to reward on the preceding trial.

3. With a sample of 40 *Ss*, it was found that a gradient after reward appeared according to position as predicted by the guessing-sequence hypothesis. The evidence indicated the lack of gradient effects in the Thorndike data and of a before gradient in the guessing-sequence findings.

4. In a free number-guessing situation, a pronounced group tendency appeared for *Ss* to respond sequentially even when they were instructed to guess randomly. More than two-thirds of the responses of 77 *Ss* fell within the limits of plus or minus two units of the preceding number. Rarely did the same number follow itself.

5. The implications of these findings are discussed and the conclusion drawn that the evidence definitely ascribes the after gradient of the 'spread of effect' to number-guessing habits.

(Manuscript received January 12, 1948)

References

1. Jenkins, W. O., & Sheffield, F. D. Rehearsal and guessing habits as sources of the 'spread of effect.' *J. exp. Psychol.*, 1946, **36**, 316–330.
2. Smith, M. H., Jr. Spread of effect and the probability bias hypothesis. Unpublished Ph.D. thesis, Stanford Univ., 1947.
3. Taylor, D. W. Spread of effect in one-trial learning in relation to the guessing-sequence hypothesis. *Amer. Psychologist*, 1947, **2**, 313. (Abstract)
4. Thorndike, E. L. A proof of the law of effect. *Science*, 1933, **77**, 173–175.
5. Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contr. Educ.*, 1935, No. 580.
6. Tilton, J. W. The effect of 'right' and 'wrong' upon the learning of nonsense syllables in multiple choice arrangement. *J. educ. Psychol.*, 1939, **30**, 95–115.
7. Tilton, J. W. Gradients of effect. *J. genet. Psychol.*, 1945, **66**, 3–19.
8. Zirkle, G. A. Success and failure in serial learning. I. The Thorndike effect. *J. exp. Psychol.*, 1946, **36**, 230–236.

DISTRIBUTED PRACTICE IN MOTOR LEARNING: PROGRESSIVELY INCREASING AND DECREASING RESTS [1]

BY BARBARA S. COOK

University of Michigan

AND

ERNEST R. HILGARD

Stanford University

The commonly found advantage of distributed over massed practice in perceptual-motor learning suggests a search for optimum conditions, and for the causal factors underlying such conditions. It is possible that the most favorable arrangements may call for a progressive change in the pattern of practice and rest as learning proceeds.

Snoddy (6) proposed that practice has a differential effect early and late on what he called primary and secondary growth. If this hypothesis is correct, distributed practice ought to be most advantageous early in learning, massed practice later in learning. In a test of this possibility, Doré and Hilgard (3), using a pursuit rotor, found the opposite to hold. That is, their subjects scored higher when rests progressively increased than when rests became progressively shorter. Snoddy (7, 8) objected to the experiment of Doré and Hilgard because the whole span of practice covered a single period. Changes of rest within a single practice session introduce interferences which, according to Snoddy, make the results meaningless as a test of his theory. He contended that if Doré and Hilgard had tested their subjects after an overnight rest they would have found no differences.

The present experiment was designed to clarify the situation. Instead of covering but one practice session, the experiment covered three daily sessions. Within each day the arrangement of practice and rest was uniform for each subject, so that interferences due to changing arrangements within the day were eliminated.[2] Also the return to practice after an overnight rest tests the residual effects of the previous practice.

Procedure

The Koerth pursuit rotor used had a ¾-in. brass target revolving at one cycle per sec. with the turntable of a phonograph. Contact with the target of the point of a hinged stylus was registered by means of an electric timer. Scores were recorded in tenths of seconds. The possible score per trial of one min. length was thus 600.

Two groups of 26 subjects, half men and women, each members of a class in introductory psychology, practiced according to the schedule of Table I. The length of each practice session was 21 min., and by the final trial of the third day both groups had had 33 trials. The difference between the groups lay in the arrangement of practice whereby Group I, with decreasing rests, began with three-min. rests and ended with 20-sec. rests, while Group II, with increasing rests,

[1] The experimentation was carried out by one of us (B. S. C.) at Stanford University during 1946–1947 in connection with a thesis for the Master of Arts degree (1). The other one of us (E. R. H.) is responsible for the report in its present form.

[2] These changes in experimental design would not have satisfied the late Dr. Snoddy, because he objected in any case to the nature of a trial in pursuit learning. The point of the experiment is not to prolong an argument over his views, but to correct some obvious shortcomings of the Doré and Hilgard experiment to which he rightly called attention.

TABLE I

ARRANGEMENT OF PRACTICE AND REST FOR THE TWO GROUPS

	Group I Decreasing Rests	Group II Increasing Rests
Day 1	6 1-minute trials, 3-minute rests	16 1-minute trials, 20-second rests
Day 2	11 1-minute trials, 1-minute rests	11 1-minute trials, 1-minute rests
Day 3	16 1-minute trials, 20-second rests	6 1-minute trials, 3-minute rests
	1 supplementary trial after a 5-minute rest	1 supplementary trial after a 5-minute rest

began with 20-sec. rests and ended with three-min. rests. A supplementary trial was run at the end of the third day following a rest of five min., in order to determine residual effects.

RESULTS

The course of performance for the two groups is plotted in Fig. 1. In common with the results of the Doré and Hilgard experiment, distribution of practice is advantageous late as well as early, within the regular practice trials. The earlier results were therefore not an artifact of practicing within a single session, for they are here confirmed over a three-day period.

The supplementary information provided by the new experiment is that after long rests performances tend to be alike, regardless of the arrangements of prior practice. This is true for overnight rests as predicted by Snoddy. It is also true over the supplementary rest of five min., where there is even a tendency toward reversal of mean scores, although the final difference is not statistically significant. This dependence of final levels of scoring on the total time elapsed from the beginning of practice, rather than upon the number or arrangement of trials, is consistent with some earlier results of Doré and Hilgard (2), and of Hilgard and Smith (5).[3]

In order to interpret the differences exhibited in Fig. 1, critical ratios have been computed for several of the comparison points. The groups were satisfactorily matched at the beginning, through assigning subjects alternately to Groups I and II (*C.R.* = 0.9). By the end of the first day, they differed significantly (*C.R* = 3.1) but they did not differ following the overnight rest (*C.R.* = 1.0). They remained alike through the second day's practice, no difference being demonstrable at the end (*C.R.* = 0.2). During the third day the mean score of the group with distributed practice was consistently superior, although on the final trial the difference was not statistically satisfactory (*C.R.* = 1.7). Even though this be interpreted as no difference, the final score has been achieved in six practice trials of the last day instead of the 16 trials of the massed practice group. Even equality of final scoring level is thus a demonstrated advantage for distributed practice, if trial-by-trial gains are considered.

On the supplementary trial following a five-min. rest at the end, scores for both groups increased, but the greatest gain was for the group whose scores had been depressed by overcrowding. The result was a reversal in mean scores, although the difference did not reach statistical significance (*C.R.* = 2.2). Because the lower scoring group made the higher scores following rest, the difference in gains was significant (*C.R.* = 4.0). It is evident that massing was con-

[3] For a discussion of these and related findings, including some recomputations of Snoddy's data, see Hilgard (4, pp. 250–257).

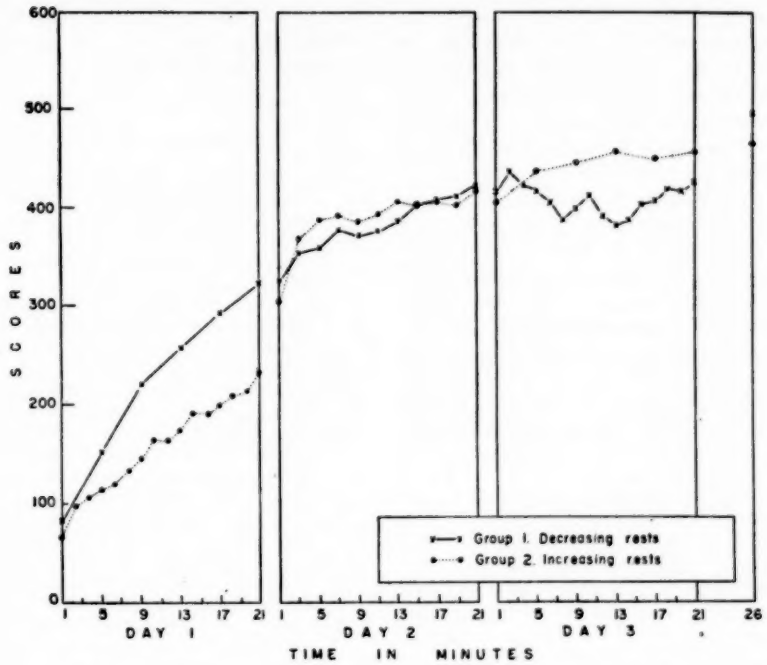


FIG. 1. Mean scores throughout practice for two groups differing in the distribution of practice and rest (N = 26 for each group)

Arrangements of practice are alike for the two groups on Day 2. On the other days the group with distributed practice makes the higher scores, whether early or late in learning. The advantage is lost, however, when both groups receive a five-min. rest at the end of practice.

tinuing to produce a depressing effect at the end of practice.

CONCLUSION

Differences at the end of practice were not significant, although one group had practiced with increasing rests, another with decreasing rests. The results do not bear out Snoddy's prediction that massing should be harmful early in practice and advantageous late in practice. They also fail to confirm the suggestion from the Doré and Hilgard experiment that there is an advantage in the reverse order of short rests early and longer rests later. Overcrowding of trials is disadvantageous throughout. It may be that the decremental effects late in learning are less serious than early in learning, but further experimentation would be necessary to determine whether or not this is so.

SUMMARY

1. Two groups of 26 subjects each practiced on the Koerth pursuit rotor during three daily 21-min. sessions. For one group the rests between one-min. trials decreased progressively from three min. on the first day to one min. on the second day to 20 sec. on

the third day. For the other group the inter-trial rests increased from 20 sec. to one min. to three min. on the successive days. By the end of the experiment, both groups thus had the same number of trials under each condition, but in reverse order.

2. Distributed practice was notably advantageous on the first day, and continued to show some advantage over massed practice on the third day. This agrees with the earlier results of Doré and Hilgard.

3. Following overnight rests, and after a five-min. rest at the end of the third day, scores of the two groups did not differ significantly, despite differences in the arrangement of previous practice.

4. The results support neither Snoddy's position that decreasing rests should be advantageous, nor Doré and Hilgard's results showing increasing rests to be advantageous. Overcrowding of trials has a decremental effect on scores both early and late in practice. Whether the effect differs early and late in learning remains to be determined.

(Manuscript received May 3, 1948)

References

1. Cook, Barbara S. *Distribution of practice early and late in learning.* M.A. thesis. On file, Stanford Univ. Library, 1947.
2. Doré, L. R., & Hilgard, E. R. Spaced practice and the maturation hypothesis. *J. Psychol.*, 1937, **4**, 245–259.
3. Doré, L. R., & Hilgard, E. R. Spaced practice as a test of Snoddy's two processes in mental growth. *J. exp. Psychol.*, 1938, **23**, 359–374.
4. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
5. Hilgard, E. R., & Smith, M. B. Distributed practice in motor learning: Score changes within and between daily sessions. *J. exp. Psychol.*, 1942, **30**, 136–146.
6. Snoddy, G. S. *Evidence for two opposed processes in mental growth.* Lancaster, Pa.: Science Press, 1935.
7. Snoddy, G. S. A reply to Doré and Hilgard. *J. exp. Psychol.*, 1938, **23**, 375–383.
8. Snoddy, G. S. Evidence for a universal shock factor in learning. *J. exp. Psychol.*, 1945, **35**, 403–417.

CONTROL OF REST INTERVAL ACTIVITIES IN EXPERIMENTS ON REMINISCENCE IN SERIAL VERBAL LEARNING[1]

BY STEPHEN WITHEY

University of Michigan

AND

CLAUDE E. BUXTON AND ALBERT ELKIN

Northwestern University

This is the second of a series of reports concerning the methodology of studies of reminiscence in verbal learning. Like the first report (3), this one is concerned with a possibly significant variation among certain experiments, especially those of Melton and Stone, and Hovland. Two different investigations are combined for purposes of this paper.

Accepting the definition of reminiscence as improvement in recall without (any) intervening practice (2), it becomes important to control the activities between learning and recall so as to prevent rehearsal. To keep these activities from producing retroactive inhibition, *E*s have usually filled the period between learning and recall with tasks which hold attention but do not require learning. The present report is concerned with the effects of color-naming activities only.

Melton and Stone, whose *S*s were "urged to name the colors as fast as possible" (6, p. 296), did not find evidence of reminiscence, whereas Hovland (4), who required the naming of colors from the memory drum at a two-sec. rate, did find it. There were other important differences between the studies, but these techniques of controlling rehearsal have never been investigated directly. And since certain investigators (1, 7) have indicated that *S*s report some rehearsal during color naming from the drum, the failure to control rehearsal might account for the greater evidence of reminiscence in the Hovland experiment.

EXPERIMENT I

Procedure

Subjects.—College students, ranging in level from freshman to graduate student, were used as *S*s. There were a total of 38, 18 women and 20 men. All were unfamiliar with nonsense-syllable experiments.

Lists.—Ten lists of syllables were prepared, six for the practice and four for the experimental sessions. Each list contained 12 syllables, typed in pica capitals. All syllables were of low association value, and were taken from an unpublished manuscript by Melton (5). The usual rules for list construction were employed, and a cue symbol served as a warning before the first item was presented. A modified Hull-type drum was used, with exposures at a two-sec. rate. There were four sec. between the last syllable and the cue symbol. (For details of lists, see Withey, 9.)

Experimental design.—The lists were learned one per day for 10 days distributed over two academic weeks, the order in which lists occurred being identical for all *S*s. At the first session, *S* read a printed page of conventional instructions silently while *E* read them aloud. Before the second and third sessions, *S* was reminded of the salient points of procedure, and his technique was corrected if he failed to follow instructions. Thereafter, instructions concerned only general reminders as to technique and any special (experimental) procedure for the day.

On the first two experimental days a two-min. rest period was interpolated as soon as *S* reached

[1] Experiment I was done by Withey for an M.A. thesis (9); Experiment II was an independent study by Elkin, with the cooperation of Prof. B. J. Underwood; the statistical analyses and preparation of this report are the work of Buxton.

a criterion of 7 out of 12 syllables correct. During one such rest period (condition 7), he was required to name color patches fastened to the tape and exposed in the regular way on the memory drum, without learning them (the technique of Hovland). At the other of these two sessions (condition 8), *S* spent the rest period naming squares of color mounted on a board (the technique of Melton and Stone). For this purpose, 50 small squares were pasted to each of four boards; the squares were aligned in rows and columns in random order. *S* was instructed to name them in any sequence he chose, but to do this at a *moderate* rate (in contrast with the rapid rate of Melton and Stone). In neither of these two conditions (which were counterbalanced for order among the *S*s) was there any mention of the supposed rehearsal-preventing character of the activities, nor any special instruction about rehearsal.

In condition 9 (for all *S*s), the rest interval was filled with naming colors from the drum, but a special instruction prior to learning urged *S* to avoid all rehearsal while naming colors. For condition 10 (for all *S*s), a special prior instruction urged *S* to avail himself of every possible opportunity to rehearse while naming colors during the rest period.

It will be noted that in the four conditions there are two sub-experiments, and the analysis of data will be presented in that fashion. Comparison of results for conditions 7 and 8 should reveal whether, without special instruction as to rehearsal, color naming from the boards or from the drum will produce differential effects at recall. Comparison of conditions 7, 9 and 10 should reveal whether, with naming from the drum constant, instructions with respect to rehearsal have any bearing on performance at recall.

It will be noted that a control (i.e., no-rest) condition was not employed, so a conventional measure of reminiscence is not available for any of the conditions employed. The control was omitted because the experiment was directly concerned with rest interval activities and with the possible effects of rehearsal, rather than with measuring reminiscence as such. In order that the experiment would afford data comparable to those of reminiscence experiments, however, all the factors known to be conducive to the appearance of reminiscence were utilized here: syllables, rather than adjectives; two-sec. rate of exposure; brief inter-list interval; serial anticipation learning; two-min. rest period; *S*s liberally practiced at shifting to rest-interval activity and back. As will be seen later (Table I), in all conditions except condition 9 the recall scores indicate that reminiscence would have been found had the usual control-condition data been utilized.

Results

Over-all practice effects.—During the experimental days and the practice days alike, *S* learned continuously to a 7/12 criterion before any variation was introduced into the task. It is therefore possible to note, day by day, the number of trials necessary to reach this criterion. Because of the fact that systematic counterbalancing was not possible during the experimental days (except for conditions 7 and 8), it is important to know whether the general curve of performance was level during the collection of experimental data. That the curve was approximately level during the four experimental sessions is shown by the daily means (with *SD*s in parentheses): 8.0 (2.4), 8.2 (3.1), 7.8 (2.2), 8.0 (2.7). Thus, although list and day variances are confounded in this design, rate of learning to the criterion on experimental lists was approximately constant.

Analysis of recall data.—Shown in Table I are unadjusted means (and *SD*s) for scores on the criterion trial, and the recall trial, for each condition. It will be noted that the criterion means are very well matched for all conditions, but that some variation appears at recall. The experimental design permits use of analysis of co-variance to test differences among

TABLE I

UNADJUSTED MEANS OF CRITERION SCORES AND RECALL SCORES OF EXPERIMENT I, AND THEIR *SD*s ($N = 38$)

	Drum, Neutral Set	Boards, Neutral Set	Drum, Set to Avoid Reh.	Drum, Set to Rehearse
$M_{cri.}$	7.34	7.39	7.47	7.39
SD	.55	.69	.72	.69
$M_{rec.}$	7.61	7.92	7.21	7.92
SD	1.32	1.35	1.20	1.55

recall means. In such an analysis (8, p. 333), the error estimate is a residual variance from which has been removed variance due to individual differences. (This is possible because the same *Ss* were used in all four conditions of the experiment.)

The question of whether color naming from the drum or from the boards has a greater effect upon recall scores (conditions 7 vs. 8) may be answered very briefly. The *F*-ratio for this variance is 1.59 (1 and 16 *df*), a value well below the .05 level of significance. Thus, the two rest interval activities are not significantly different in their effects upon recall, when no special set with respect to rehearsal has been given. Further consideration will be given later to the effects of rapid color naming.

Conditions 7–9–10 may be analyzed to see whether special instructions about rehearsal during color naming from the drum have any influence upon recall scores. As the unadjusted recall means of Table I suggest, set not to rehearse, no special set, and set to rehearse, produce increasing recall scores, in that order. The *F*-ratio for this variance (7.73; $df = 1$ and 73; .01 level = 6.98) indicates, beyond the .01 level of confidence, that the variable of instruction with respect to rehearsal is significant. Since, in most experiments employing naming from the drum, this variable has not been explicitly controlled, it can only be concluded that if *Ss* instruct *themselves* at random an added source of error variance is present.

Summary of Experiment I.—(1) The fact that Melton and Stone did not obtain reminiscence, whereas Hovland did, is probably not to be attributed to the difference in the activity with which their rest intervals were filled. This may be inferred from the present finding that the two activities do not produce significantly different recall scores when all major variables other than type of activity are constant. (2) Failure to instruct *Ss* with respect to rehearsal, in experiments employing color naming from the drum, probably will result in variance attributable to self-instruction by *S*. The possibility of this is indicated by a significant variation in recall scores when instructions with respect to rehearsal are the variable.

Experiment II

Problem.—At the time Experiment I was designed, it was thought desirable to compare color naming from the drum and from the boards when only the type of task, and not the 'pressure' on *S*, was varied. When it was found that the two rest-interval activities did not differentially affect recall, a further check on rate of color naming from boards was performed. It will be recalled that Melton and Stone (6) urged their *Ss* to name colors as rapidly as possible. This could well lead to blocking and other signs of emotionality and this in turn to lessened recall scores.

Procedure

Subjects.—All *Ss* were students in an undergraduate experimental psychology course. There were 19 women and one man.

Lists.—Adjective lists taken from Melton's materials (5) were used for serial learning. There were three 12 item lists, one for a practice day and two for experimental sessions. The order of lists was constant for all *Ss*. A Harvard-type drum was used, with exposures at a 2.9-sec. rate, and with 15 sec. between the last syllable and the cue symbol.

Experimental design.—On the practice day all *Ss* were treated alike, being given conventional instructions for serial learning, being corrected on learning technique as necessary, etc. When the criterion of 7 out of 12 responses correct was reached, *S* was given two min. of color naming from boards, like those of Experiment I, and

then a recall trial. Colors were named in time with a metronome beating at 66 counts per min.

The next two sessions were like the first, except that the experimental variable was introduced by requiring different rates of color naming during the rest intervals. The slow rate was set by a metronome beat of 74 per min. and the fast by a beat of 132. The order of slow and fast conditions was counterbalanced. That the two rates of color naming were different was shown in part by the fact that no *S* had any difficulty in keeping the slow pace, whereas 17 of 20 *S*s showed some blocking under the fast rate, as well as signs of minor stress.

Results

The principal outcome of the experiment may be seen in Table II.

TABLE II

UNADJUSTED MEANS OF CRITERION SCORES AND RECALL SCORES OF EXPERIMENT II, AND THEIR *SD*s ($N = 20$)

	Fast Rate	Slow Rate
$M_{cri.}$	7.50	8.00
SD	.92	.90
$M_{rec.}$	7.15	7.70
SD	1.62	2.03

On the criterion trial there was a slight superiority of scores for the condition where slow color-naming was to be introduced. On the recall trial for both conditions there was some loss in mean score, as was the case in Melton and Stone's data, with the slight superiority of the slow-rate condition being increased somewhat. Analysis of co-variance of the type used for Experiment I showed that there was practically no difference between the adjusted recall means. In other words, it does not appear that rate of color naming from the boards is a significant factor in the later attempted recall of adjectives. This finding supplements that of Experiment I, by suggesting that if a rapid rate of color naming had been used in Experiment I, condition 8, the results would not have been changed.

(Manuscript received April 7, 1948)

REFERENCES

1. BAKANOFSKY, M. B. The comparative effectiveness of two learning techniques, correction and non-correction, for the appearance of reminiscence in the serial anticipation learning of nonsense syllables. M.A. Thesis, State Univ. of Iowa, 1945.
2. BUXTON, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, **40**, 313–340.
3. BUXTON, C. E., & ROSS, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
4. HOVLAND, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
5. MELTON, A. W. Materials for use in experimental studies of the learning and retention of verbal habits. (Unpublished manuscript, 1940)
6. MELTON, A. W., & STONE, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
7. ROSS, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. M.A. Thesis, State Univ. of Iowa, 1945.
8. SNEDECOR, G. W. *Statistical methods.* (4th ed.) Ames: Iowa State College Press, 1946.
9. WITHEY, S. B. Relationship between reminiscence effects and rest interval activity in serial anticipation learning. M.A. Thesis, Northwestern Univ., 1947.

AN EXPERIMENTAL ANALYSIS OF LATENT LEARNING [1]

BY JOHN P. SEWARD

University of California at Los Angeles

Problem

How does a rat learn a maze? According to reinforcement theory (8), when a rat chooses the correct path his responses lead more directly to food and thus receive a greater share of primary or secondary reinforcement than when he enters a blind alley. In later trials he reacts on the basis of the differential strengths of habit thus produced. According to sign-Gestalt theory (25), a rat learns a maze by running it; he then chooses on the basis of anticipated consequences. If hungry he takes the path that leads, in his experience, to a food place. To decide between these views we must separate running the maze from feeding in it; learning should then be retarded by the first view but not by the second. The latent learning experiment is thus a crucial one. In it the rat first explores an empty maze; he then finds food at some point in it; whereupon he has been found to run the maze almost as well as if fed there from the beginning (1, 26).

Elsewhere I have tried to envisage the essential mechanism of maze learning as revealed by the latent learning experiment (21). It involves two stages: (1) the rat builds up differential anticipatory responses to the right and left cues at the choice point; (2) one of these responses becomes associated with food. As a result the rat now reacts to the appropriate cue because response to it is 'reinforced' (i.e., facilitated) by food anticipation. To test this interpretation let us suppose that a rat is allowed to explore a simple T-maze with all outside cues removed and with endboxes clearly differentiated but out of sight from the choice point. He is then put directly into one of the endboxes, where he finds and eats food for the first time. After this experience he is placed in the starting box and allowed to choose either arm of the maze. The above interpretation leads us to predict that the rat will go directly to the box in which he has been fed. There is one important proviso: the endboxes must be distinguished in such a way as to arouse *different responses* which can be partially reinstated at the choice point on the critical trial. Otherwise the rat will be left with no medium to serve as a connection between choice-point cues and anticipated food.

To test these deductions the following experiments were performed.

Experiments

Main Experiment

Method.—An alley maze was built in the form of a T, with removable endboxes projecting at right angles to the crossbar on the same side as the stem. Dimensions were as follows: stem, 4 ft.; crossbar, 8 ft. overall, 4 ft. on each side; endboxes, 16 in. Alleys and endboxes were 4 in. wide and 5½ in. high. Sides and floor were of wood, painted gray inside except for one pair of endboxes; the roof was of hardware cloth. The entire maze was covered with cheesecloth draped over wires 18 in. above the maze and hanging down on both sides of the alleys. Four 15-watt lamps were mounted on uprights at both ends of the stem and crossbar of the T; when these lights were turned on in the darkened room the cheesecloth became a one-way screen that effectively eliminated extra-maze visual cues. The endboxes and the first 10 in. of the stem could be closed off at will by manually operated sliding doors.

[1] Aided by a grant from the Committee on Research of the University of California. I wish to thank Miss Susan Nail and Mr. Nissim Levy for their expert assistance in collecting a large part of the data.

An attempt was made to delimit the kinds of cue that can mediate 'reinforcement' under these conditions. Experiments on delayed response (17, 19, 28, 29) and on discrimination with delayed reward (5, 18) and secondary generalization (6) led me to believe that non-verbal animals would be unable to anticipate colors. There is evidence, however (5, 15, 17, 18, 19), that motor responses can better serve a symbolic function (22). Two pairs of endboxes were therefore used. One pair had wooden floors and were painted white and black, respectively. The other pair were painted gray, but one had a floor of hardware cloth and the other a floor of sponge rubber. Otherwise all endboxes were alike. Each one was provided with a small food cup that was always present in a fixed position, whether filled or empty. My prediction was that rats run to white and black boxes and then fed in one of them would be at a loss to know which way to turn for food. Rats run to differential floors, however, would have proprioceptive cues at their disposal that would determine a correct choice.

Thirty-two hooded rats of both sexes, ranging in age from 5 to 6½ months, made up the experimental group. Sixteen of these were used with the white and black boxes (group WB), 16 with the hard and soft floors (group HS). They were run in groups of 10 or 12 at a time, half to one pair of boxes and half to the other. The white and hard-floor boxes were always at the right, the black and soft-floor boxes at the left.

The rats were fed one hour a day at 24-hour intervals. Their diet was a modified Steenbock ration in the form of a mash, which they ate in a feeding cage at the end of each day's experimenting. After two or three days on this schedule, they were given six days of preliminary adaptation in a straight alley 68 in. long that resembled the maze in every other respect. On the first three days they were put in the alley in groups of five or six and allowed to explore it for two periods of 10 or 15 min. each. During one period they found one endbox (e.g., white or hard) at the end of the alley; in the second period they found the other (black or soft). On the last three days the rats were placed individually in the starting box of the alley and allowed to run once to each of the two endboxes assigned to their group. They were now considered ready for the experiment.

On three days the rats were put in the maze five or six at a time and allowed to explore it for 30 min. a day. All doors were open, giving access to the endboxes, which of course contained no food. A rough count was kept of the number of times per day each rat entered each endbox. Recorded entrances, which underestimate the actual totals, ranged from 2 to 13 with a mean of 6.8.

On the fourth day the procedure for a given rat was as follows: Three min. of solitary exploration of the maze were followed by an isolation period of about 25 min. (range: 17–34) in a remote detention box. The rat was then put directly into one endbox, through the roof, where for the first time he found the door closed and the food cup filled with his accustomed food. (Half of the rats were assigned at random to each endbox.) Here he was kept for one min. or, if eating was delayed, until he had settled down to eat. At that point he was lifted out and put down in the starting box, the doors were raised, and he was given a free choice. Food odors were controlled by placing a large cup of mash just outside of each endbox. The approximate time interval was recorded between the rat's removal from the endbox and his arrival at the choice point. The distribution of these intervals yielded the following statistics: mean, 31.7 sec.; *SD*, 13.7 sec.; median, 29.5 sec.; *Q*, 5 sec.

Results.—Of the 32 rats, 28, or 87.5 percent, went directly from the choice point to the endbox in which they had been fed. The difference between this figure and chance has a significance ratio of 4.2, which refutes the null hypothesis at a confidence level of .01. Of the 28 correct, 15 belonged to group HS, 13 to group WB.

An error was counted if the rat definitely left the choice point in the direction opposite to food. This criterion is arbitrary. Of the four errors, one was made by a rat in group WB that started down the wrong alley, stopped within a foot of the choice point, nosed a piece of cheesecloth accidentally caught under the roof, then turned and took the other path. If this performance were not counted an error, the percent correct would be raised to 91.

Interpretation.—The success of group HS was predicted from a hypothesis as to the nature of latent learning in a T-maze. This hypothesis stated that, during exploration, differential anticipatory responses

come to be aroused at the choice point and that, if one of these is then associated with food, it will serve to facilitate the corresponding choice. The success of group WB was not predicted, however, for it necessitated the implausible assumption that rats can anticipate the whiteness or blackness of an endbox.

The results of an experiment are crucial for a hypothesis to the extent that other explanations are ruled out. Although it is seldom possible to rule out every alternative, at least the more obvious ones should be controlled. To be at all definitive for the present hypothesis a test must fulfill a minimum of three conditions: (1) a correct choice on the test trial must depend on the preceding exploration; (2) it must depend on cues not present at the choice point; (3) it must depend on the association of one set of these cues with food. In order to determine whether the above test met these conditions, three control experiments were conducted.

Control Experiment 1

Problem.—Our first requirement was to see if rats without previous familiarization with the maze could solve the problem of finding the endbox in which they had just been fed. It is conceivable that they could, either on the basis of differential stimuli common to endboxes and choice point or on the basis of vestibular cues produced by being carried from endbox to starting box.

Method.—Rats were put through the same course of adaptation in the straight alley as the experimental group; i.e., group exploration, 10 min. a day to each endbox for three days, followed by one individual run a day to each endbox for three days. On the seventh day the test procedure of the earlier group was duplicated. There was no preliminary exploration of the maze; each rat was put into one or the other endbox and fed briefly, then put in the starting box for his test. Thus, the only difference in procedure was that the control rats had never been in the maze before the test trial.

Three groups of rats were treated in this way, about half of each group running to each pair of endboxes. Group A, composed of 25 male and female hooded rats 3½ to 4 months old, was considered unsatisfactory in the light of Maier's finding (12) that young rats were inferior to those over four months old in his three-table 'reasoning' test, of which the present experiment may be considered a variation. Group B, consisting of 11 albino rats between six and seven months old, was also inadequate because of the inferior vision of albino compared with pigmented rats (20). The 19 hooded rats of group C, six to seven months old, constituted an acceptable control.

Results.—The results of all three groups appear in Table I. Of the 27 rats in group WB, 12 chose correctly; in group HS the number correct was 15 out of 28.

TABLE I

CORRECT CHOICES IN CONTROL GROUPS UNFAMILIAR WITH MAZE

Group	*N*	Correct Choices	
		No.	%
A	25	14	56
B	11	6	55
C	19	7	37
Total	55	27	49

It was possible that the difference between the experimental and control results might be due to a longer delay interval for the controls between removal from endbox and arrival at choice point. That this was not the case was shown by the following statistics: mean, 26.2 sec.; *SD*, 9.6 sec.; median, 23.1 sec.; *Q*, 4.8 sec.

Interpretation.—The only obvious objection to this experiment is that the very novelty of the maze may have prevented the rats from making use of cues otherwise available. After six days of preliminary adaptation in the straight runway, however, the rats

gave no sign of unusual disturbance on the test trial. The conclusion seems justified that prior exploration of the maze was indispensable to successful performance.

Control Experiment 2

Problem.—Our second requirement was to see if prior exploration had actually resulted in 'knowledge' of the endboxes and their distinguishing features. One other possibility suggested itself. A small section under the cheesecloth, containing a shaded lamp mounted on a gray-painted upright, was visible through the roof of each endbox and also, at a much greater distance and a 90° angle, from the choice point. It is conceivable that during maze exploration these two views of some differential cue might have acquired stimulus equivalence for the animal and thus, between them, mediated a choice in the direction of food. To check this possibility a control experiment was conducted in two parts, A and B. In part A all differential cues from inside the endboxes were removed; in part B all differential cues *except* those inside the endboxes were removed.

Part A. *Method.*—Two new endboxes were made, as nearly alike as possible, with walls and floor of wood painted a uniform gray. Otherwise, they were identical with those previously used. The procedure was also the same except that after each occasion in the maze the endboxes were interchanged. That is, they were interchanged after the first, second, and third days of group exploration and again after the three-min. individual exploration on the fourth day. Intra-endbox stimuli could thus give no dependable clue to the location of food.

Results.—Twelve male hooded rats five to six months old were run under these conditions. Their score was exactly 50 percent. Their failure was not due to inadequate exploration, since during the first three days their recorded entrances averaged 6.2 per endbox per day, with a range from 3 to 12. Nor was it due to the length of interval between removal from endbox and arrival at choice point; the mean delay was 30.6 sec., the median 27.5 sec. Despite the small number of animals, I saw little use in prolonging this part of the experiment.

Interpretation.—It is true that this method of removing endbox cues also introduced contradictory ones that might be partly responsible for the chance results. But, since all other cues were constant, we can at least conclude that these remaining cues were not the only ones that the successful rats in the first experiment had used. Our next task was to remove them to see if they played any part at all.

Part B. *Method.*—Both original pairs of endboxes were used and the procedure was the same as that of the experimental group, with a single exception. On the day of the test, after the preliminary exploration period, the assigned endbox was moved from its place in the maze and, still pointing in the same direction, was put just behind the starting box, where it formed an extension of the stem of the T. Lighting and one-way screening were similar to those in the original positions but were identical for both endboxes. The rat was then placed in the endbox in its new location and, after eating there, was put in the starting box for his critical trial.

A total of 48 rats took part in this experiment, divided equally into groups WB and HS. One difficulty was that putting the endbox back in position lengthened the time after eating before the rat could be released to make his choice. This defect was not corrected until 36 hooded rats (group D), aged 5½ to 8 months, had been run. Their mean delay interval was 41.7 sec. (median, 40 sec.; *SD*, 7.6 sec.), representing an increase of about 10 sec. over the experimental group. At that point a complete set of duplicate boxes was made, to be used solely for replacements on the critical trial. Twelve more rats (group E) were now run with a mean delay of 18.8 sec. and a median of 20 sec. These rats were albinos, six aged 4–4½ months, six aged 6–6½ months, assigned to groups WB and HS, respectively.

TABLE II

CORRECT CHOICES IN CONTROL GROUPS WITH PREFEEDING IN NEUTRAL LOCATION

Group	N	Correct Choices	
		No.	%
D	36	19	53
E	12	7	58
Total	48	26	54

Results.—The results, presented separately for groups D and E, appear in Table II. Both before and after the change they approximated a chance level.

Special interest attaches to a breakdown of these results according to type of endbox, since here if anywhere, with other cues eliminated, a difference between visual and proprioceptive symbols might be expected to appear. Table III shows the results separately for group WB, run to white and black endboxes, and for group HS, run to hard and soft floors. Neither score deviates significantly from chance.

Interpretation.—It is unlikely that the failure of group E was due to their lack of pigmentation, since five out of the seven correct choices occurred in group WB, where such lack should be a greater handicap. The fact that group E failed although the interval between prefeeding and choice was shorter than in the experimental group strongly implies that group D's failure was not due to their lengthened delay. Maze exploration, moreover, was adequate, as indicated by a mean number of daily entrances of 8.4 per endbox, with a range from 2 to 20.

TABLE III

RESULTS OF TABLE II REGROUPED ACCORDING TO MODE OF DIFFERENTIATION OF ENDBOXES

Group	N	Correct Choices	
		No.	%
WB	24	11	46
HS	24	15	63

One possible factor in the failure of these groups is that feeding the rats behind the starting box may have substituted spatial cues to the new location for the ones that were removed. To a rat, food-in-a-white-box-just-behind-me may not mean food-in-the-same-white-box-some-distance-to-the-northeast at all. I noted that one rat tried to climb the back wall of the starting box and regret not having made systematic observations of such behavior, though it certainly did not impress me with its frequency. What we can conclude, however, is that intra-endbox cues were not exclusively relied on by the successful performers in this situation.

How are we to interpret the seemingly paradoxical results of parts A and B? The rats in our experiment apparently based their choice on a number of cues, of which three possible ones have been mentioned: (1) differences inside the endboxes, (2) objects (lamps, uprights) under the one-way screen visible at the end of each alley, (3) vestibular impulses. To these cues may be added a fourth; viz., the symmetrical arrangement of the lighted alley-ends as seen at right angles from inside the endboxes. Rats will presumably use all cues available in a problem situation. When all were present, as in the experimental group, the animals were able to solve the problem. But when one or more of them were eliminated, as in the control groups, they could not do so. A plausible conclusion is that several cues, *including those in the endboxes*, were necessary to correct performance. Logically, in fact, this

conclusion is unavoidable. Let us consider the three possible alternatives: (1) The rats reacted to cues inside the endboxes alone. (2) The rats reacted to cues outside the endboxes alone. (3) The rats reacted to cues both inside and outside the endboxes. We now see that the results of part A are compatible with (1) and (3) but not with (2) and that the results of part B are compatible with (2) and (3) but not with (1). Therefore (3) alone can be accepted. It follows that successful performance in our experimental situation does depend in part on differential anticipatory reactions.

What can we conclude from the lack of difference between groups WB and HS as to the relative efficacy of visual and tactual-kinesthetic cues in mediating 'anticipation'? Unfortunately, nothing. Although we know the rats made use of cues received in the endboxes, we have no way of telling what those cues were. Aside from those intentionally provided by the experimenter there may have been others both inside the boxes—e.g., the odor, grain, or surface irregularities of the wood—and outside but visible therefrom—viz., those parts of the apparatus within the one-way screen. Our secondary assumption, that anticipatory responses must be motor, is thus neither confirmed nor refuted.

Control Experiment 3

Problem.—The first two control experiments indicated that the experimental group had built up differential anticipatory responses to endbox cues. One question still remains: did their correct choices actually depend on the food they had found in one endbox? There is an alternative explanation. In criticizing Maier's three-table 'reasoning' test, to which ours is obviously similar, Hull (7) proposed that Maier's rats learned a discrimination based on the perseverative trace from the table last visited. Since my rats were given only one test, they had no chance to learn such a discrimination. Nevertheless the perseverative trace from one endbox may have been enough in itself to determine a choice in the appropriate direction even without the food incentive. One way to test this possibility would be to repeat the first experiment but omit the food. A more convincing test of the importance of food, however, would be to leave the food in and see if the rats could find it when endbox traces were equalized. The second method was accordingly adopted.

Method.—Thirty-two hooded rats were used, 20 males and 12 females, 4½ to 6 months old. The first nine rats were divided between the original two pairs of endboxes, but since no clear-cut difference between WB and HS groups had appeared so far, that part of the problem was now abandoned. For the next 10 rats the white and black boxes were used, but the white box had a strip of hardware cloth over the floor and the floor of the black box was covered with a piece of glass. For the last 13 rats the original 'hard' and 'soft' boxes were used, but a rubber mat was put over the sponge rubber in the 'soft' box to keep it from being chewed.

The procedure before the test day was the same as in the first experiment. On that day the individual exploration period was increased from three to five min. The important change, however, concerned the pretest procedure. Each rat was placed for two min. in one endbox, then for two min. in the other; the test followed immediately. Each rat found food in only one box. Half of the rats were fed in the right-hand box, half in the left; within each of these groups half were put in the food box first, half in the empty box first. Perseverative traces from the two endboxes were thus equalized for the group as a whole, leaving 'knowledge of the location of food' as the only basis for better-than-chance performance.

Results.—Of the 32 rats 26, or 81.3 percent, chose the food side.[2] This

[2] As in the main experiment, there was one doubtful case. This rat went half-way down

figure differs from chance with a t of 3.53, which is significant at the .01 level. Though slightly lower than the score in the main experiment, it is not significantly so. The present finding therefore has a twofold value: (1) it confirms the ability of rats to solve the problem set in the original experiment; (2) it shows that that ability depended not merely on a perseverative endbox trace but on an association of one endbox with food.

Discussion

A proposed mechanism of latent learning.—Our primary hypothesis sprang from an earlier attempt to come to closer grips with the process of latent learning (**21**). The support given by the present results may justify a brief but in some respects more explicit restatement of the basic formulation.

Associative learning is handled through the concept of the *surrogate response*, defined as follows:

Definition 1: When a stimulus (S) evokes a response (R) it does so by activating a central process, to be called a surrogate response ($\overline{rs}$), which has a reaction potential to that R.[3]

The usefulness of this concept as a central counterpart of the "response-produced stimulus" (**16**) has been demonstrated elsewhere (**21, 22**). Its role in learning is embodied in the following proposition:

Postulate 1: If two Ss evoke Rs in close temporal proximity the $\overline{rs}$ of S_1-R_1 acquires a functional connection with $\overline{rs}$ of S_2-R_2. ("Functional connection" is equivalent to Hull's "habit strength" (**8**).)

From the above definition and postulate we may derive the fundamental theorem of conditioning:

Theorem 1: A S to R_1, if closely followed by R_2, acquires a functional connection with R_2.

Selective 'learning' is deduced with the help of Hull's concepts of *drive* and *reaction potential* (**8**, p. 238 ff.). For our purpose a part of his postulate 7 (**8**, p. 253) may be recast as follows: *A functional connection between S and R is energized by a primary drive into reaction potential ($_SE_R$).* We may also borrow from Thorndike the concept of the *satisfier* (S_G) (**24**), equivalent to Hull's 'reinforcing agent' or 'incentive' (**8**, p. 131). We are now in a position to state a principle from which, combined with others, the selection of successful R's may be deduced:

Postulate 2: Reaction potential between S and R ($_SE_R$) is increased by the simultaneous activation of the $\overline{rs}$ to a satisfier ($\overline{rs}_G$).

Postulate 2 is a revision of Thorndike's law of effect and of Hull's doctrine of reinforcement. It differs from their statements chiefly by making satisfiers influence response-evocation rather than habit strength, performance rather than learning. According to Thorndike (**24**) the successful R wins out because a satisfier sets off a confirming reaction that automatically strengthens any S-R connections active at the time. Hull, too (**8**, p. 178), supposes that need reduction results in an increment of habit strength between accompanying S's and R's. In the present version when S-R is followed by a satisfier (S_G-R_G) a functional connection is formed between their surrogates (postulate 1). When S recurs in the presence of a drive it therefore activates $\overline{rs}_G$, which increases $_SE_R$ (postulate 2), and thus brings about the selection of R. Such an effect of $\overline{rs}_G$ on $_SE_R$ might also be aptly called *reinforcement*, if the word's relation to habit strength were not so firmly fixed.

Of the two postulates, only the first is truly a learning principle. 'Selective

---

the wrong alley, then doubled back and proceeded to the food box. A more lenient criterion of correctness would raise the percentage to 84.4.

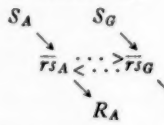
[3] To be sufficiently comprehensive, 'response' must be taken to include sense organ adjustments, slight changes of postural tonus, and the like. In practice these must often be inferred rather than observed, although such procedures must always be considered tentative.

learning,' in the view here taken, is a misnomer, since the behavior it denotes can be explained as a result of associative learning combined with the motivating effects of drives and satisfiers.

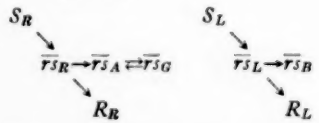
To see how these principles work out in practice, we may apply them to the behavior observed in the present experiment. Let us represent a functional connection in the process of formation by $\cdots >$ and one already formed by $\rightarrow$. Let S_R be the differential stimulus pattern received by a rat at the choice point as he faces right, S_L as he faces left. Let S_A be the differential stimulus pattern he meets in the right endbox, S_B in the left. Let S_G stand for food and let the rat be fed in the right endbox. We assume without complicating the diagrams that the animal is hungry and the functional connections are adequately energized. The operation of the above principles under the conditions described for the experimental group may then be represented in the following series of diagrams:

1. Maze exploration

2. Prefeeding



3. Test



The first two diagrams show the essential connections formed during the daily exploration periods and while eating in the endbox on the fourth day (postulate 1). The third diagram presents the competing reaction tendencies on the test trial as the rat faces right and left at the choice point. The arousal of $\overline{rs}_G$ by S_R but not by S_L (except through stimulus generalization) is responsible for a stronger reaction potential to R_R (postulate 2).

Evidence from related experiments.—The above fragment of a 'miniature system' is intended as a step toward a conceptual framework for the phenomena of associative learning and selection in general. Its value is most obvious, however, in deducing responses that depend, as in the present experiment, on a combination of separately formed habits. Besides latent learning experiments of the conventional maze type, the two most closely related designs are found in Maier's 'reasoning' tests (**11**, **13**) and in studies of so-called 'secondary generalization' (**8**).

As already indicated, Maier's experimental arrangement is so similar to mine as to be its prototype. The chief difference lies in the fact that my set-up largely excluded stimuli common to feeding place and choice point, whereas Maier did not try to control this factor. In addition, my rats had only one test each, whereas Maier was concerned with the ability of his rats to vary their choices from day to day according to the place of prefeeding. A third departure occurred in control experiment 3 when my rats were given pretest experience in *both* endboxes, rather than only one. But the pattern of the two experiments is the same. It would not be surprising, therefore, to find that the interpretation here proposed could readily be extended to include his findings.

In secondary generalization two responses (R_1, R_2) are repeatedly evoked in one situation (S_1) and two responses (R_2, R_3) in another (S_2); whereupon S_1 is found to evoke R_2 without further training. Many variations on this theme have been reported, from Brogden's 'sensory preconditioning' (**2**) and Graham's 'transfer of conditioning' (**4**) in dogs to the 'indirect conditioning' procedure of Shipley (**23**) and Lumsdaine (**10**) with humans. The supposed common factor R_2 has been aroused by food

(3, 27), by shock (14), or by the endbox of an alley (15). Informal attempts at qualitative deduction of these findings suggest that they can be readily handled at this level by the present postulate set. Its advantage lies in the availability of an intermediate reaction, the *surrogate response* $\overline{rs}_2$, to bridge the gap between S_1 and R_3.

Certain of these experiments in their turn give some tentative support to the implication of postulate 1 and theorem 1 that indirect conditioning may take place through central connections without the overt occurrence of the intermediate response. Lumsdaine (10) found that S_1 occasionally evoked R_3 simultaneously with or even before R_2. Graham (4) verified this finding, and reported that R_3 frequently occurred without R_2, although incipient reactions may have gone unnoticed. Loucks (9) found that rabbits continued to give R_3 in response to S_1 after proprioceptive cues from R_2 had been abolished by a spinal anesthetic. Such results hardly 'prove' the validity of the surrogate-response concept, but they do lend it a degree of plausibility.

Summary

An experiment was devised to test a hypothesis about the essential requirements for latent learning at a choice point. The hypothesis was that these consist of two separable factors: (1) associations between choice-point stimuli and alternative consequences of going right or left, giving rise to differential anticipatory reactions at the choice point; (2) association between one of these consequences and a satisfier. To test this hypothesis extra-maze cues were largely excluded from a simple alley T-maze. Rats were allowed to explore without food; they were then fed in one endbox and immediately thereafter given a free choice in the maze. Twenty-eight out of 32 rats, or 87.5 percent, chose the path to the endbox in which they had been fed.

Three control experiments were run in an attempt to delimit more precisely the factors involved in this performance:

1. A control group prefed and tested without previous experience in the maze failed to exceed chance.

2. To isolate the critical cues operating on the test trial, all differences between the right and left endboxes were eliminated for one control group; all cues *outside* the endboxes were removed for a second group. In neither case did the score exceed chance. These results indicated that successful performance depended in part on differential anticipatory reactions to endbox cues.

3. The main experiment was repeated with fresh rats, but they were placed in both endboxes, one with and one without food, just before their free choice. Twenty-six of 32 rats, or 81.4 percent, chose correctly, thus establishing the role of the food incentive.

A secondary hypothesis was that to give rise to differential anticipatory reactions at the choice point the consequences must have distinguishable motor components. To test this hypothesis half of the rats in the experimental and first two control experiments were run to one black and one white endbox, half to a pair of endboxes with floors of hardware cloth and sponge rubber, respectively. Since results with the two pairs of endboxes did not differ significantly, they failed to confirm the hypothesis.

Two postulates were proposed as basic principles of association and selection; they were applied to the results of the present experiment and their relationship to experiments on 'reasoning' and 'secondary generalization' was pointed out.

(Manuscript received May 10, 1948)

References

1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
2. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
3. Estes, W. K. Discriminative conditioning. I. A discriminative property of conditioned anticipation. *J. exp. Psychol.*, 1943, **32**, 150–155.
4. Graham, D. T. Experimental transfer of conditioning in dogs. *J. exp. Psychol.*, 1944, **34**, 486–493.
5. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
6. Honzik, C. H., & Tolman, E. C. The perception of spatial relations by the rat: a type of response not easily explained by conditioning. *J. comp. Psychol.*, 1936, **22**, 287–318.
7. Hull, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, **42**, 219–245.
8. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
9. Loucks, R. B. An experimental investigation of chained responses in indirect conditioning. *Psychol. Bull.*, 1939, **36**, 575–576.
10. Lumsdaine, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reactions. *Psychol. Bull.*, 1939, **36**, 650. Reported in: Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940, 230–232.
11. Maier, N. R. F. Reasoning in white rats *Comp. Psychol. Monogr.*, 1929, **6**, No. 3.
12. Maier, N. R. F. Age and intelligence in rats. *J. comp. Psychol.*, 1932, **13**, 1–6.
13. Maier, N. R. F. The effect of cerebral destruction on reasoning and learning in rats. *J. comp. Neurol.*, 1932, **54**, 45–75.
14. May, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.
15. Miller, N. E. A reply to "Sign-Gestalt or conditioned reflex?". *Psychol. Rev.*, 1935, **42**, 280–292.
16. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
17. Nissen, H. W., Riesen, A. H., & Nowlis, V. Delayed response and discrimination learning by chimpanzees. *J. comp. Psychol.*, 1938, **26**, 361–386.
18. Riesen, A. H. Delayed reward in discrimination learning by chimpanzees. *Comp. Psychol. Monogr.*, 1940, **15**, No. 5.
19. Riesen, A. H., & Nissen, H. W. Nonspatial delayed response by the matching technique. *J. comp. Psychol.*, 1942, **34**, 307–313.
20. Robinson, E. W., & Wever, E. G. Visual distance perception in the rat. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 233–239.
21. Seward, J. P. A theoretical derivation of latent learning. *Psychol. Rev.*, 1947, **54**, 83–98.
22. Seward, J. P. The sign of a symbol: a reply to Professor Allport. *Psychol. Rev.*, 1948, **55**, 277–296.
23. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.
24. Thorndike, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton-Century, 1935.
25. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
26. Tolman, E. C., & Honzik, C. H. Introduction and removal of reward and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
27. Walker, K. C. The effect of a discriminative stimulus transferred to a previously unassociated response. *J. exp. Psychol.*, 1942, **31**, 312–321.
28. Yerkes, R. M., & Nissen, H. W. Prelinguistic sign behavior in chimpanzee. *Science*, 1939, **89**, 585–587.
29. Yerkes, R. M., & Yerkes, D. N. Concerning memory in the chimpanzee. *J. comp. Psychol.*, 1928, **8**, 237–271.

CONTINUITY IN HUMAN LEARNING

BY W. C. H. PRENTICE

Swarthmore College

The 'continuity' *vs.* 'non-continuity' issue has usually been confined to discrimination learning in the laboratory rat. The original observations by Lashley (**7**), the ingenious experiment by McCulloch and Pratt (**13**), the countering experiments by Krechevsky (**6**), the series of investigations by Spence (**16, 17, 19**), and all the controversial articles and notes and comments (e.g., **1, 9, 11, 12, 15**) have concentrated their attention on what happens when an albino or hooded rat is confronted with a problem requiring sensory discrimination.

It seems likely, however, that most of these authors are not primarily concerned with the behavior of the rat *qua* rat. With the possible exception of Spence, who has rather special views about the nature of the scientific process (see **18**), most of those who have been involved in this prolonged controversy would perhaps be willing to entertain the notion that the rat's behavior has some implications for a theory of human learning. The pedagogical importance of the assumption that human learning is continuous (or non-continuous) might be great in certain cases; but even more clearly, the assumption of continuity and the assumption of non-continuity lead to quite different kinds of thinking about central nervous mechanisms underlying learning. If these are the issues with which we are ultimately to be concerned, it will sometime be necessary to come to grips with the problem of continuity at the human level.

In addition, it has previously (**12, 15**) been pointed out that in some sense a crucial experiment is impossible at the animal level. Lacking any but inferential evidence concerning the existence of 'hypotheses' or the end of the 'pre-solution period,' one must always be uneasy concerning the rigor with which specific experiments have demonstrated either 'continuity' (see, for example, Krechevsky's (**6**) criticisms of the original McCulloch and Pratt paper) or 'non-continuity' (see, for example, Spence's (**17**) objections to Krechevsky's technique).

To assume that we are in any better case when dealing with human subjects is to assume that we can, as a matter of fact, have access to certain kinds of evidence for such constructs as 'hypothesis,' kinds of evidence that are lacking when we observe rats. This assumption is, no doubt, debatable. In a previous paper (**14**), the present author has argued that the phenomena of human experience can be ignored as scientific data only by a thoroughgoing dualist who believes that in some way 'mind' and 'body' are different substances and cannot interact. But even if that point be granted, it is still true that one must make inferences about the experiences of 'the other one' and that human phenomena are not directly available to an experimenter in the way that behavior samples are. The issue between rats and men must rest then upon the confidence of a particular experimenter in his inferences; does he trust the verbal reports of adult human subjects more or less than he trusts runnings and jumpings of rats? Which is the more reliable information when one wants to infer the state of a problem-solution?

In the experiment to be reported here, the working assumption has been that the state of problem-solution may be inferred with more accuracy from the verbal reports of the subjects than from their choice behavior alone. Supposing that assumption to be justified, what should the typical reversal-test for continuous learning show? If two groups of human subjects are both taught the same problem to the same criterion, but if one of them has previously been trained on precisely the opposite problem, what will be the difference in their performance on the common task? More particularly, how will their scores compare in the case when one group has been transferred from the reverse problem *before* he has either shown any systematic response to the relevant cues or indicated by his verbal reactions that he is aware of them?

Since the 'continuity' principle holds that learning is essentially a matter of strengthening connections between specific stimuli and specific responses, and that the course of solution is determined rigorously by a summation of the rewards and punishments administered trial by trial, that principle should lead us to expect that *O* will have built up *some* tendency to make the 'wrong' [1] response by the time he is 'reversed.' That tendency will presumably have to be overcome before *O* is again in the neutral state with respect to the two choices that he was in at the beginning; and only then can he begin to build up response tendencies that favor the 'correct' choice. Assume, for instance, that *O* has initially equal tendencies to choose 'circle' and 'square' in a given stimulus situation. If now circle is rewarded or square punished every time that stimulus situation reappears, *O* must increase his tendency to choose circle, *even if he has not correctly grasped the critical aspects of the stimulus situation which make the circle correct.* In other words, the continuity principle seems to demand that when human subjects are given preliminary practice on a problem which is exactly the opposite of one to be learned later, the later learning should be slowed up, even when the change to the reversed problem occurs before the *O*s have begun to attend to relevant parts of the situation, as indicated by their verbal reports.

The hypothesis that learning may be 'non-continuous,' however, is based on the assumption that factors of attention, perceptual organization, and the like will limit the efficacy of reward and punishment to something less than the entire stimulus situation. Associations will then be formed only between the responses and those aspects of the stimulus situation which are in some sense obtrusive or central or 'attended to.' In other words, an *O* who has been making his choices on the basis of cues that have no relation to the true solution, and who has not yet noticed the existence of those really relevant cues, might be quite unaffected by reversal. So far as he is concerned, this may be the same problem he had before; nothing appears to have changed, and he proceeds to the solution at exactly the same rate he would have if no change had been introduced. The changed conditions are purely extrinsic to *O*'s problem and so will not affect his responses. They are changes made behind the scenes, as it were, since his attention to other things has been such as to make him oblivious of them. The 'non-continuity' hypothesis is that associations will not be formed between responses and such 'behind-

[1] 'Wrong' and 'correct' refer to the final problem which is, of course, the precise opposite of the one on which *O* is originally trained.

the-scenes' stimuli. If the *O*'s free verbal account of his thinking process has been accurate and complete, and if it has omitted all mention of the relevant cues, one might suppose that reversal should make no difference.

Procedure

Group 1.—*O* sat at a table facing a fiber-board panel in the center of which was a hinged screen that could be raised or lowered by *E*. Above the screen was a small light. On the table below the screen were two telegraph keys, one marked with a circle and one marked with a square. *E* read the following instructions:

> Behind the screen are two windows. Whenever I raise the screen you will see two designs, one in each of the windows. One of them will always be a circle and the other will always be a square, but in other respects they will differ from time to time. On each exposure, you are to choose *either* the circle or the square and signal your choice by pushing the proper key. If you decide to choose the circle, push the key marked with a circle; if you choose the square, push the key marked with a square. If you choose the *correct* one, the light at the top of the panel will go on; if you choose the *wrong* one, you will hear a buzzer. Your task is to learn to get the light every time, never the buzzer. Just one more thing; in order that I may keep track of how your learning proceeds, please try to *think aloud*. If you find this embarrassing or unnatural at first, at least make sure that you tell me why you made the choice you did on each trial. Are there any questions?
>
> There is no hurry. On each trial the screen will remain open until you have made a choice. Just try to reach perfect scores in as few *trials* as possible. Ready? Here is the first pair.

The materials were eight cardboard squares presented in pairs to *O*. Each card was a simple drawing of a single square or circle; two of the circles were black on a white background and two were white on a dark gray background; one of each such pair was large (three in. in diameter) and one was small (1¾ in. in diameter); the squares were similarly large-white, small-white, large-black, and small-black. The pairs were so arranged that a choice was always available between two sizes, two colors, and two shapes; in addition, each pair was presented equally often in each left-right arrangement, so that position was another variable. The switch controlling the light and buzzer was manipulated by *E* so that 'circle' would always be correct when any figure with a dark background was on *O*'s right; 'square' would, of course, be wrong in these cases and correct in all others. This arrangement produced 50 percent success if other principles of choice were used, e.g., 'smaller is correct,' 'large squares are correct,' 'alternation: square, circle, square, circle, etc.'

With one class of exceptions, each *O* continued until a criterion of 12 successive correct choices had been reached. The exceptions were those *O*s who, at the end of 100 trials, were still making only chance successes and had shown no indication by their verbal reports that they were 'on the right track.' These *O*s will be referred to below as 'Failures.'

Group 2.—Group 2 was treated precisely like Group 1 with one exception. Group 2 spend its first *20 trials* hearing the buzzer for choices which were 'correct' throughout for Group 1 and seeing the light for choices that were 'wrong' for Group 1. After 20 trials, these *O*s were 'reversed'; that is, the lights and buzzers now started following the pattern used for Group 1 throughout their sessions. This means that Group 2 spent 20 trials being trained on exactly the opposite task from the one which they were later to learn. Of course *O* was given no indication that this was not all one problem or that any change of procedure was involved.

All *O*s were undergraduates in the Johns Hopkins University; they were naive concerning the 'continuity' problem and had never served in learning experiments before. About two-thirds of them were men and one-third women. No special attempt was made to equate the two groups, but *O*s were assigned to them at random. The learning session with each *O* occupied from 30 min. to one hour, about 20 sec. elapsing between each presentation and no time limit being put on *O* in making the discrimination.

Some sample protocols will illustrate the kinds of 'hypothesis' expressed by various *O*s at different stages of the solution. (See p. 190.)

The asterisks in the first three cases indicate the first trial on which any sign of attention to the proper combination of variables appeared. In the case of Observer *Sch.*, this 'hypothesis' appeared before reversal, as it did in eight other *O*s. The original instance of a correct hypothesis often preceded complete solution by many trials even for *O*s in the control (or non-reversal) group. The case of Observer *Sea.* is included to illustrate the random and repetitive nature of the solutions suggested by those who never succeeded in solving the problem. In most, but not all, such cases, *O* appeared to become fixated on the notion of temporal sequences.

One cannot rigorously rule out the possibility of 'multiple hypotheses,' the possibility, that is,

O	Trial No.	Comment
Fr.	1	Just guessing.
	4	Circle, because it is the smaller.
	13	I think squares have generally been correct on the left hand side.
	27*	The circle on a gray background is correct on the right side.
Ri.	4	Circle, because it has a dark background.
	5	Square because it has a dark background.
	6	That theory was wrong, just guessing now.
	15	I am choosing the large figure on white cards now, but I think there is some (temporal) pattern.
	33*	Circles on white cards on the left are correct.
Sch.	2	Trying circles until I see when they are correct.
	3	Think it has something to do with position.
	7	I suspect there is some order of lefts and rights.
	9	I am going to stop trying to memorize the order.
	10	I just noticed the color difference.
	12*	Testing to see whether the white circle on gray on the left is always correct.
Sea.	2	Still just guessing.
	3	Circle on white was correct last time so I am choosing the square which is now on white.
	4	That was wrong; just guessing.
	6	Alternation: circle, square, etc.
	18	Card with small hole in the center.
	23	Circle, circle, square, circle, circle, square, etc.
	24	Have to start over, just guessing.
	99	I am sure it is some kind of sequence.

that *O* might be reporting only one of several relationships to which he was, as a matter of fact, attending. In general, however, the fluency of the observers was such that it seems unlikely that we have misjudged the time of earliest appearance of the proper hypothesis by more than a few trials. Post-experimental interviews confirm the generalization that the solution typically appeared as a tentative hypothesis based on summarizing the results of the preceding three or four trials. All *O*s reported that they were unable to recall much about what had occurred more than a few trials back, and the tendency of all but a very few *O*s to repeat errors is consistent with this report.

Results

The main findings can be stated very simply. Group 1, the control group, learned to make errorless runs in just over 41 trials. Group 2, which had been taught a reversed form for 20 trials, learned the regular problem in a total of just over 62 trials. Table I gives the standard deviations and critical ratio. There is good reason to accept this result as indicating that the experimental group was slowed down by the reversal treatment.

TABLE I

Means and *SD*'s of Trials by Groups 1 and 2 to an Errorless Run, and C.R. of the Difference Between the Means

Group	*N*	Trials	*SD*	σ_m
1	20	41.37	17.96	4.23
2	20	62.42	26.37	6.20

Diff. = 21.05 $\sigma_{Diff.}$ = 7.5 *C.R.* = 2.81

Since all *O*s in group 2 were reversed at exactly the 21st trial, those who exceeded the average rate of learning showed signs of recognizing the relevant relationships *before* reversal. To check the possibility that these subjects were contributing a bias to the total result reported in Table I, all *O*s in both groups who had given overt evidence of attending to the correct relationships before the twentieth trial were omitted, and the remaining *O*s were considered separately. The result, of course, is to remove the better learners and to raise all averages correspondingly. Group 1 in its diminished form averages 50.73 trials, while Group 2 averages 73.50. Table II gives the treatment in full. It seems likely that the differences be-

TABLE II

COMPARISON OF GROUPS 1 AND 2 AFTER REMOVAL OF *O*s SHOWING INSIGHT BEFORE PROBLEM REVERSAL

Group	*N*	Trials	*SD*	σ_m
1	12	50.73	17.78	5.62
2	12	73.50	23.48	7.40

Diff. = 22.77 $\sigma_{Diff.}$ = 9.28 *C.R.* = 2.46

tween the groups must still be considered significant.

These results conform on their face to the predictions made from 'continuity' principles and fail to fit the kind of prediction that stems from 'non-continuity' assumptions. There are, however, some complexities to the data.

The reader should notice that *O*s in Group 2 can be considered slower learners than those in Group 1 *only if we include all their trials*, or in other words *only if we consider the regular problem and its reversed form to be all one learning task.*

If we consider the number of trials taken by Group 2 *after* beginning the basic form of the problem, it will be seen to be almost exactly the same as the number taken by Group 1. That is an important fact, since, as Krechevsky (6) has pointed out, under the implications of the continuity principle this whole experimental design may be considered as a problem in transfer of training. In such a view, the question would be whether training on problem 1 (first 20 trials) affected learning on problem 2 (the basic form). The continuity hypothesis assumes that the strength of a stimulus-response tendency is a continuous function of reinforcement. It must then predict that *O*s whose tendencies to choose 'square' have been rewarded during the first 20 trials must have those tendencies overcome by opposite reinforcements in order to make them equal (in likelihood of choosing circle or squate) to the control group. One might expect that it should take 20 or so trials to overcome the effect of the first 20 and that only then could the experimental group be put on an equal footing with the controls. That would mean that the differences in Tables I and II would be of the order of *40* trials instead of 20. To put it another way, we should find negative transfer in the sense that the experimental group should need significantly more trials *after reversal* than the control group needed all together. This result does not appear, and consequently the experiment must cause difficulty for the kind of associationism that has been fostered by (e.g.) Spence.

On the other hand, the 'non-continuity' view does not profit much from such an analysis. For on the assumptions of that position, the present experiment consisted not of two problems, but of one. So long as *O* is not paying attention to the relevant variables, E's changing the order of rewards and punishments has no psychological significance. Everything else remains the same; if changing the reinforcement pattern makes this in some sense a different problem, it must be either (a) because *O* has really noticed the relevant variables (which contradicts our experimental assumptions) or (b) because the continuity view is really correct and associations have been made even though *O* was not 'aware' of the stimuli that were participating in the associations. To accept (b) is, of course, to give up the whole controversy. The possibility that (a) is true seems very slight. If it is, of course, the present experimental results are simply irrelevant. But *no subject* was aware of any change in procedure; this fact is established both by reference to their

trial-by-trial protocols and by direct questioning after the experiment ended. For all *O*s the first 20 trials and the rest of the learning were concerned with the same problem having, at all times, only one implied solution.

Rejecting reservations (a) and (b), as we must, the non-continuity hypothesis must predict *no difference* between Groups 1 and 2. So long as *O* has not begun *in any way* to react systematically to the to-be-reversed cues, the first 20 trials for Group 1 are just like the first 20 trials for Group 2. They consist, presumably, of trying out and discarding certain hypotheses or response patterns that are equally unrewarding for both our 'reversed' and 'normal' conditions. Only after those patterns have been given up, can *O* turn to others and begin (in both Group 1 and Group 2) to learn about the *then* correct reaction to the real cues.

The whole non-continuity position rests on the assumption that rewards and punishments will be effective only for stimuli that are being responded to in an organized fashion. If those stimulus conditions whose reinforcement is reversed have not yet been responded to in this sense, then surely the first 20 trials and the later ones must be considered psychologically part of a single task. *That* task should be solved just as fast by the experimental *O*s as by the controls, i.e., *total* trials should be the same. This is not the case. And, as a consequence, the non-continuity theory is given as much trouble by the discrepancy between a predicted zero and the existing twenty trials of difference between our groups as the continuity view is by the discrepancy between the predicted forty and the existing twenty.

A further, and independent, complexity arises when we notice that those *O*s who 'failed' (as defined above) might be expected on 'continuity' grounds to come primarily from Group 2. In fact, there were 13 failures in Group 1 and 18 in Group 2, a result that is in the predicted direction but probably is not great enough to be significant. Using the assumption that 20 passes and 15 failures (the mean for both groups) is the chance expectation, two distributions of 'passes' and 'failures' as different as these would occur by chance approximately half the time.

Finally, neither group shows any sign that the number of successful choices made (presumably by chance) during the first 20 trials is in any way associated with the number of trials required to reach criterion. On the assumption that learning is continuous, these zero correlations would only occur if the effectiveness of our 'rewards' and 'punishments' were exactly equal. The result is more simply explained by non-continuity assumptions.

Discussion

The present experiment can certainly not be considered as support for the 'non-continuity' hypothesis. In fact, to some extent it may be considered as an example of the failure of 'non-continuity' principles even under the most favorable of circumstances. For it had been argued (1, 15) that the artificial behavioral criterion of Krechevsky must always mark the end of the pre-solution period too late, and these arguments had implied that if evidence of 'hypothesis-formation' could be got as soon as it appeared, the results of the McCulloch-Pratt and later investigations would be upset. In the present case, every effort was made to include in the experimental group only those *O*s who had given objective (verbal) evidence of

their failure to attend to the relevant discriminanda. By using human subjects, moreover, the experiment insured that 'hypotheses' in some relatively familiar sense were indeed involved in the learning process. Nevertheless, it appears that early associative connections are in some way formed without the intervention of systematic activities of the learner.

On the other hand, one can hardly hold that learning is a 'continuous' process in the sense in which that term has sometimes been used. The experiments of Haire (3), McClelland (10), and Harlow (4), for instance, have shown that many of the phenomena of learning are quite inconsistent with the proposal that the strength of a learned tendency is continuously increased by each association between a response and a reinforcement or between a stimulus and a reinforcement. The latter two authors, it is true, have leaned over backward to state that their results are not contradictory of Spence's notions, but there is no real reason to believe that the 'continuity' principle is of much use if it holds only with respect to one kind of subject in one stage of one experimental situation. We are thus confronted with the fact that sequences of stimulus-response-reward appear to summate their influences under certain conditions (including those of this experiment), but that under other conditions the frequency with which such a sequence has appeared may be irrelevant to the determination of the final level of behavior. The Lashley-Krechevsky assumption that associations are formed only among those aspects of the stimulus-situation that are immediately 'being attended to' seems not to be true. And yet the view that *O* is somehow 'forced' to respond in the way that has led to success the largest number of times in the past is just as clearly unsatisfactory when stated without reservation.

The plain fact is that the relationship between cognition and action in complex learning situations still needs to be elucidated. For nearly twenty years, psychologists have been concerned with the distinction between learning and performance, but we still have inadequate operational tools for defining the learning independently of the performance, and in at least some instances (most of the 'continuity' experiments) it seems easier to predict action without the postulation of some mediating cognition than with it.

The importance of the present investigation lies in its implication that some such mechanical principle is useful in describing certain aspects even of adult learning which is clearly the result of ego-involved striving. It is probably true that such usefulness is confined to the early stages of the learning, since McClelland (10) has shown that intentional reversal of responses occurs with ease once a discrimination has been mastered. It must be emphasized, however, that the cognitive theories of learning have not shown their capacity to account for this kind of result even in early stages.

Summary

A conditional visual discrimination was taught to two groups of human subjects. One group learned the problem to criterion after having previously had 20 trials on the opposite problem. All *O*s gave running verbal accounts of their reasoning. Even when all *O*s who had showed any signs of attending to the relevant cues before reversal were dropped from consideration, this group learned significantly *slower* than a control group. These results seem to support a 'con-

tinuity' notion. On the other hand, the decrement in rate of learning was almost exactly equal to the number of trials on the 'reversed' problem. This means that there is, strictly speaking, no negative transfer; so that certain difficulties are raised for interpretations based on simple associative concepts. It is also true that no significant difference appeared between the total number of *O*s in the two groups who failed ever to learn the correct solution, and that there is no correlation between number of successful choices in early trials and number of trials to reach criterion. These results are interpreted as indicating the need for experimental and theoretical advances toward an understanding of the relation between theories which emphasize respectively the cognitive and the response aspects of learning.

(Manuscript received May 4, 1948)

References

1. Haire, M. A note concerning McCulloch's discussion of discrimination habits. *Psychol. Rev.*, 1939, **46**, 298–303.
2. Haire, M. Some experimental data relevant to field and associative theories of discriminative learning. *J. gen. Psychol.*, 1942, **26**, 267–288.
3. Haire, M. The role or repetition in discriminative learning. (Abstract) in: Harvard University *Summaries of Theses 1942*, Cambridge: Harvard Univ., 1946.
4. Harlow, H. F. Studies in discrimination learning by monkeys: I. The learning of discrimination series and the reversal of discrimination series. *J. gen. Psychol.*, 1944, **30**, 3–12.
5. Krechevsky, I. A note concerning the nature of discrimination learning in animals. *Psychol. Rev.*, 1937, **44**, 97–104.
6. Krechevsky, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, **45**, 107–133.
7. Lashley, K. S. *Brain mechanisms and intelligence*. Chicago: Univ. Chicago Press, 1929.
8. Lashley, K. S. An examination of the "continuity" theory as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26**, 241–265.
9. Lashley, K. S., & Wade, Marjorie. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
10. McClelland, D. C. Studies in serial verbal discrimination learning. IV. Habit reversal after two degrees of learning. *J. exp. Psychol.*, 1943, **33**, 457–470.
11. McCulloch, T. L. Comment on the formation of discrimination habits. *Psychol. Rev.*, 1939, **46**, 75–85.
12. McCulloch, T. L. Reply to a note on discrimination habits. *Psychol. Rev.*, 1939, **46**, 304–307.
13. McCulloch, T. L., & Pratt, J. C. A study of the pre-solution period in weight discrimination by white rats. *J. comp. Psychol.*, 1934, **18**, 271–290.
14. Prentice, W. C. H. The interruption of tasks. *Psychol. Rev.*, 1944, **51**, 329–340.
15. Prentice, W. C. H. Operationism and psychological theory: a note. *Psychol. Rev.*, 1946, **53**, 247–249.
16. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
17. Spence, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.
18. Spence, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.
19. Spence, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.

SENSITIZATION OF THE BETA-RESPONSE AS A FUNCTION OF THE WAVELENGTH OF THE STIMULUS[1]

BY DAVID A. GRANT, JOHN P. HORNSETH, AND HAROLD W. HAKE

University of Wisconsin

The object of this investigation was to determine to what degree sensitization of the β-response of the human eyelid is a function of the wavelength of the stimulus. The β-response has been defined as a secondary eyelid response to light which occurs at a latency between 120 and 240 msec. (3). During dark adaptation the β-response increases in frequency and magnitude (4, 7), but this 'sensitization' effect has not been observed when dark adaptation has been prevented (4, 5, 8, 9). With respect to modifiability under instructed set, the β-response has appeared to be more like a reflex than like a learned response (7, 10) although the mechanism involved is still largely unknown.

In the present experiment the dark adapting *S* was presented stimuli of different wavelengths in order to determine whether short wavelengths of light or long wavelengths of light were more effective in eliciting larger and more frequent β-responses. It was thus possible to observe whether the behavior of the β-response was consistent with the hypothesis that frequency and magnitude increases of this response during dark adaptation were brought about by the scotopic adaptation mechanism.[2]

[1] This research was supported in part by the Graduate Research Committee of the University of Wisconsin from special funds voted by the State Legislature for 1946-47.

[2] The general experimental approach was patterned after that of Adrian (1), who showed that a secondary component of the human electro-retinogram was intimately related to sensitization of the rods of the retina.

Apparatus and Procedure

The general procedure was the same as that followed in the recent studies cited above. Stimuli were presented and responses were recorded by a modified Dodge photochronograph.

The test stimuli consisted of light from a photoflood lamp which was passed through a color filter and a circular, milk-glass window 10 cm. in diameter, situated 60 cm. directly in front of the *S*. The duration of each stimulus was about 750 msec. Five different stimuli were obtained by using five of the Wratten '70-series' filters (Nos. 70, 71a, 72, 73, and 75).[3] As pointed out by Adrian (1, p. 86) the '70-series' filters pass narrow bands of wavelengths, and furthermore, the amount of light transmitted by the different filters was such as to make the stimuli appear approximately equal in brightness. The brightness of the stimuli was of the order of magnitude of 140–200 millilamberts as measured by a Macbeth illuminometer.[4]

With each *S* the same general procedure was followed:

1. Light adaptation (1500 millilamberts) two min.
2. Dark adaptation (.01 millilamberts) one min.
3. *Test Series No. 1* (10 test stimuli)
4. Dark adaptation (.01 millilamberts) five min.
5. *Test Series No. 2* (10 test stimuli)
6. Dark adaptation (.01 millilamberts) five min.
7. *Test Series No. 3* (10 test stimuli)
8. Dark adaptation (.01 millilamberts) five min.
9. *Test Series No. 4* (10 test stimuli)

[3] Manufactured by the Eastman Kodak Company, Rochester, New York. Transmission curves for these filters are given by Adrian (1) and in *Wratten Light Filters*, Eastman Kodak Co., 1940.

[4] Exact equation of stimulus brightness was, of course, impossible to obtain in this experiment. An equation during photopic condition of the retina would not hold for scotopic vision and vice versa. For this reason, only the general order of magnitude is given.

The *S* was instructed to fixate a spot at the center of the milk-glass window at all times during the experiment. When no test stimuli were being presented, the brightness of this window remained at .01 millilamberts to permit a considerable degree of dark adaptation. During each Test Series the five test stimuli were presented twice in a double fatigue order. The precise order was varied from series to series and from *S* to *S* so as to balance out ordinal effects. Within each series the temporal interval between stimuli was varied randomly from 35 to 55 sec.

Forty men and women students from the class in elementary psychology served as the *S*s in the present study.

Results

The average magnitude of β-responses during each test series is presented graphically in Fig. 1 for each of the five stimuli. Fig. 1 shows clearly that the average magnitude of β-responses increases for all stimuli during the course of dark adaptation, but the most important feature of Fig. 1 is the fact that the greatest increments take place for filters 73 and 75, which transmit light from the region of the spectrum which contains the short wavelengths. The increments for long-wavelength stimuli, filters 70, 71a and 72, though clear-cut, are not as great as for the short wavelength stimuli.

The analysis of variance of the magnitude data of Fig. 1 is presented in Table I. The experimental design

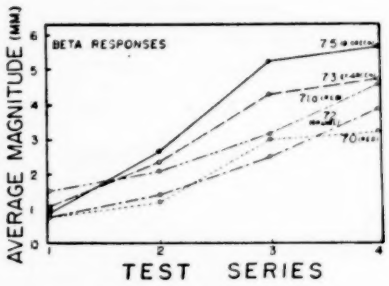


Fig. 1. Average magnitude of β-responses plotted against successive Test Series with wavelength of the stimulus as a parameter

can be considered to be a factorial design (11) in which four independent variables are mutually orthogonal to (or independent of) each other. The four independent variables are *Stimuli*, *Test Series*, *Subjects*, and *Order*. Of these factors, the first three are self-explanatory, but the last requires a word of explanation. It arises because each stimulus was presented twice in each Test Series, permitting a difference to occur between the responses evoked by the first and second presentation of each stimulus in the various Test Series. Such a difference is referred to here as an *Order* effect. The four-factor (third order) interaction mean square of 14.53 based upon 468 degrees of freedom is used as the error estimate; it is the denominator term for all of the *F*'s in Table I.

The general impressions gained from Fig. 1 are borne out by the analysis in Table I. The differences between stimuli are highly significant, as are the differences between Test Series. Order shows no effect, but there are significant differences between individual *S*s in the magnitude of the β-responses. The combined effects or interactions are interesting but not of great importance to this study. The statistical significance of the *F* for Subject-Test Series interaction means simply that the rate and degree of increase in magnitude of β-response differed from *S* to *S*. The marginally significant *F* for the Stimulus-Subject-Test Series interaction suggests that the above-mentioned individual differences in rate of increment also varied with respect to the stimulus. One marginally significant effect, the Order-Subject-Test Series interaction, is almost impossible to explain except as a chance 'error of the first kind' (11). The relationship between the average magnitude of the

TABLE I

ANALYSIS OF VARIANCE OF THE MAGNITUDE OF BETA RESPONSES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	*F*
Filter	4	538.69	134.67	9.27**
Order	1	9.76	9.76	1/1.49
Subject	39	23,531.07	603.36	41.52**
Test Series	3	2,877.55	959.18	66.01**
Filter × Order	4	105.46	26.36	1.81
Filter × Subject	156	2,607.36	16.71	1.15
Filter × Test Series	12	283.04	23.59	1.62
Order × Subject	39	505.06	12.95	1/1.12
Order × Test Series	3	107.15	35.71	2.46
Subject × Test Series	117	10,601.83	90.61	6.24**
Filter × Order × Subject	156	2,630.34	16.86	1.16
Filter × Order × Test Series	12	97.47	8.12	1/1.79
Filter × Subject × Test Series	468	9,232.71	19.73	1.36*
Order × Subject × Test Series	117	2,776.73	23.73	1.63*
Filter × Order × Subject × Test Series (Error)	468	6,800.53	14.53	—
Total	1599	62,704.75	—	—

* — 5% confidence level
** — .1% confidence level

β-response and the wavelength of the stimulus is shown in Fig. 2. The corresponding functions are also given for the α-response or primary reflex to light (latency 60–100 msec.) and the longer latency responses (latency greater than 240 msec.) which, for the purpose of convenience only, are called γ-responses. It will be noted that the general trend of the magnitude of β-response is upward with the shorter wavelengths of light while the general trend of the magnitude of α-responses and γ-responses is downward.

A statistical test of the significance of the upward trend of the magnitude of the β-response as a function of wavelength presents some difficulties because the scale for wavelength is not necessarily linear with respect to retinal function. Further difficulties arise because the filters are not spaced at precisely equal intervals along the wavelength scale. For the purpose of an approximate test, however, it was assumed that the filters were equally spaced, and an Alexander trend analysis (2) was made of average magnitude as a function of the five stimuli. The trend analysis resulted in a highly significant F ($P < .001$) for *group slope* and for *group deviation from linearity*. This indicated that the linear part of the upward trend of Fig. 2 was significant, and that there was a further non-linear component. The trend test also yielded significant F's *between individual slopes* and *be-*

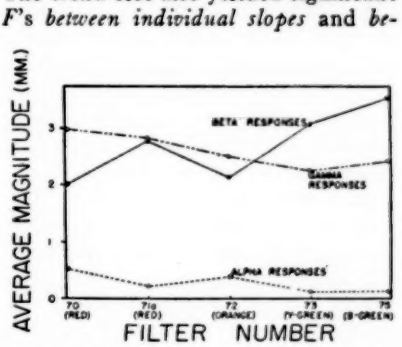


FIG. 2. Average magnitude of α-, β-, and γ-responses plotted as a function of wavelength of the stimulus

tween individual means, which were in line with the results of the analysis in Table I.

In discussing Fig. 2 it was noted that the average magnitude of α-responses and γ-responses decreased as the wavelength of the stimulus was decreased. The frequency of α- and γ-responses was not sufficiently great in this experiment to justify elaborate statistical analysis, but neither response showed a clear-cut sensitization trend during the experiment. The γ-responses increased in frequency and magnitude during the first five min. of dark adaptation, but the trends thereafter were irregular. If anything, the α-response and γ-response data suggest that these eyelid closures may be mediated primarily by the cones.

Discussion

The findings of this experiment were rather clear. The previously reported increases in frequency and magnitude of the human β-response were again found during dark adaptation. These increments were observed to be greatest for stimuli of short wavelengths.

The data support the hypothesis that the sensitization of the β-response is brought about by the increase in rod sensitivity which occurs during dark adaptation. It should be clear that although the results obtained are precisely those which would be predicted from the hypothesis, the hypothesis is by no means proved. If the sensitization effect is caused peripherally by rod adaptation, the other functional differences between the β-response and the primary reflex (4, 7) remain still to be explained. As an example, it might be pointed out that the difference in the latencies of the α-response and β-response is of an order of magnitude quite different from that existing between the latencies of the cones and the rods (1).

Summary

The degree of sensitization of the β-response or secondary eyelid response to light during dark adaptation was studied as a function of the wavelength of the stimulus. This was done by testing 40 human *Ss* with five stimuli of different wavelengths at various stages of dark adaptation. The experiment was a four factor design in which the data could be treated by an analysis of variance. The principal finding and conclusion were as follows:

1. The frequency and magnitude of β-responses increased during dark adaptation, and the increases were greatest for stimuli of short wavelengths.

2. This result supports the hypothesis that sensitization of the β-response is brought about by the increase in sensitivity of the rods which takes place during dark adaptation.

(Manuscript received May 13, 1948)

References

1. Adrian, E. D. The electric response of the human eye. *J. Physiol.*, 1945, **104**, 84–104.
2. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
3. Grant, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, **35**, 393–402.
4. Grant, D. A., & Norris, Eugenia B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.
5. Grant, D. A., & Norris, Eugenia B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–433.
6. Grant, D. A., Norris, Eugenia B., & Boissard, Suzanne. Dark adaptation and the pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1947, **37**, 434–439.

7. GRANT, D. A., NORRIS, EUGENIA B., & HORNSETH, J. P. Sensitization of the beta-response as affected by verbally induced set and counter-reinforcement. *Amer. J. Psychol.*, 1948, **61**, 66–72.

8. GRANT, D. A., HAKE, H. W., & SCHNEIDER, DOROTHY E. Effects of pre-testing with the conditioned stimulus upon the extinction of the conditioned eyelid response. *Amer. J. Psychol.*, 1948, **61**, 243–246.

9. GRANT, D. A., & SCHNEIDER, DOROTHY E. Intensity of the conditioned stimulus and the strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1949, **39**, 690–696.

10. NORRIS, EUGENIA B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter-reinforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.

11. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1946.

THE EFFECT OF INSTRUCTIONS UPON SENSORY PRE-CONDITIONING OF HUMAN SUBJECTS *

BY RUBE CHERNIKOFF AND W. J. BROGDEN

University of Wisconsin

INTRODUCTION

In the report of a recent experiment by Brogden (3), a procedure is described by means of which the phenomenon of sensory pre-conditioning was obtained with human *Ss*. A reaction time situation, with depression of a telegraph key to light, provided the response mechanism. *Ss* of the experimental group were presented with 10 trials of a light and tone while they waited for *E* to adjust the apparatus for a reaction time experiment. After instructions for reaction time and 30 practice trials of depressing the key to light, the critical post-test of 10 trials of tone alone was given. The *Ss* of the experimental group gave a significantly greater frequency of response to the tone in the post-test than did *Ss* in control groups who were given the same treatment except that the pre-training treatment involved the presentation of no stimuli in one case (control for sensory generalization) and the presentation of 10 trials of tone alone in the other (control for non-contiguous presentation of light and tone). Although positive evidence of sensory pre-conditioning was obtained, the magnitude of the effect was relatively small, both in terms of the number of *Ss* in the experimental group who showed the sensory pre-conditioning effect and in the frequency of response which they gave. The original experiment on sensory pre-conditioning in dogs (1) showed the effect to be of greater magnitude and a recent experiment by Karn (4) on sensory pre-conditioning of human *Ss* also resulted in a relatively greater magnitude of sensory pre-conditioning. These latter two experiments differ from the one described above at least in that the response mechanism was established by conditioning procedures and not by instructions.

The present experiment was designed to test several possibilities by means of which the magnitude of sensory pre-conditioning might be increased. In addition to the original set of instructions, two new sets were added. One set directed *S* not to respond to the tone in the post-test. The other set of instructions informed *S* that the light would be replaced by tone and that he was to respond or not, whichever seemed natural. For each of these instructions, the pre-training treatment of tone and light might be expected to result in a greater magnitude of response to the post-test trials of tone. The design of the experiment thus permits analysis of the magnitude of sensory pre-conditioning as a function of three different sets of instructions.

The other possibilities had to do with changes in apparatus. In the previous experiment, the source of the tone was a two-inch loud speaker in the same cabinet which contained the light. Thus, the speaker was always in the visual field of *S*. Many *Ss* in the control group for sensory generalization interpreted the presence of the loud speaker as evidence that a sound stimulus would be given, to which *E* would expect a response. It appeared

* Supported in part by the Research Committee of the Graduate School from special funds voted by the State Legislature.

possible to reduce the amount of sensory generalization by having only the light in the cabinet and by locating speakers beneath the table so that a tone of indefinite localization could be presented to *S*. It was also assumed that indefinite localization of tone might contribute to the strength of any sensory conditioning of the light and tone. In addition to this change, electronic control apparatus and an electronic chronoscope were constructed to provide more precise control of the stimulus and more precise recording of latency of the key pressing response (reaction time). The effect of these latter changes may be assessed by comparing the results for *Ss* in the present experiment who received the original instructions with the results for the original experiment on sensory pre-conditioning of human *Ss* (3).

Experimental Procedure

Throughout the experiment, *S* sat at a table in a sound-insulated, light-shielded room. A small metal cabinet with a sloping front was fastened to the table directly in front of *S*. This cabinet contained a circular hole two in. in diameter in the center of the sloping section which was covered by a milk glass plate that could be illuminated by a seven-watt light bulb. A telegraph key, securely attached to the table, was mounted to the right of the cabinet. A pair of two-in. loud speakers was fastened beneath the table, one on each side. Since the right and left speakers were an equal distance from *S*'s head, the sound stimulus was of indeterminate localization. A pair of polished silver electrodes with a rubber wrist strap was connected to a plug on the left side of the cabinet on the table. Cables from the cabinet, the speakers and key were connected to apparatus in the adjacent control room. In the wall of the experimental room facing *S* was a two-foot square window containing an x-ray mirror and two plates of clear glass of different thickness. This functioned as a one-way visual screen when illumination of the control room was less than that in the experimental room.

The rest of the apparatus was located in the control room, from which point all stimuli could be presented, data recorded, and *Ss* observed. The stimuli used were illumination of the milk glass plate in the cabinet, and activation of the speakers under the table by a complex tone of low frequency. Measurements of reflection from the milk glass plate of the cabinet were made with a Macbeth Illuminometer at the locus of *S*'s head—approximately 16 in. from the cabinet. Readings were obtained of 9.5 millilamberts under general room illumination and of 155.2 millilamberts when the bulb behind the glass plate was activated. The intensity of the tone was measured at the approximate locus of *S*'s head by a General Radio Sound Level Meter, Type 759-B, with the weighting control set at the C position. Tone intensity was adjusted to a meter reading of 43 db above 10^{-18} watts/sq. cm. This duplicated the intensity of stimulus at *S*'s head which was used in the previous experiment, even though the location of the sound source was changed from a single speaker in the cabinet of the light source to twin speakers under the table. Thus, the magnitude of both tone and light were the same as in previous experiments (2, 3) for which their equality had been obtained by the judgments of four observers.

In the previous instrumental set-up a Dunlap chronoscope, with electromagnetic clutches operated by batteries, was used. Timing of the stimuli was obtained by a synchronous motor timer which closed relay circuits for pre-set time intervals. Because of the variability found in the operation of the magnetic clutches, and the synchronous-motor-relay system, an electronic timing and chronoscope system was constructed.

The electronic chronoscope consists of a 1000 cycle audio-oscillator of the Wien bridge type (7) and a three decade electronic counter similar to the commercially available Potter Electronic Counters (5). The input stage of the counter changes a sine-wave signal into square wave pulses of the same frequency. Subsequent stages and the neon lamp recording system of the counter provide a count of the total number of pulses transmitted up to 1000. For greater counts, the instrument repeats itself in cycles of 1000. The counter is reliable up to counting rates of 20,000/sec. With an input signal from the 1000 cycle oscillator, the counter reads directly in msec. The accuracy of counter readings in msec. depends primarily on the accuracy of the 1000 cycle signal. Any deviation in frequency from 1000 cycles results in a similar deviation in the measurement of time in msec. Because changes in line voltage and in temperature produce changes in oscillator frequency, a continuous method of calibration was provided to maintain the signal to the counter exactly at 1000 cycles. Station WWV, operated by the National Bureau of Standards in Washington,

transmits audio frequency signals of 1440 and 4000 cycles per sec. that are accurate to one part in 50,000,000. A short wave radio receiver was tuned to this station and the signal was filtered so that only the 4000 cycle component remained. This signal was fed to the horizontal plates of a cathode ray oscilloscope, the vertical plates of which were connected to the 1000 cycle oscillator. A vernier control on the 1000 cycle oscillator was adjusted so that a motionless 4–1 Lissajous figure was obtained on the scope screen. This provided counter readings in msec. with a maximum accuracy of one part in 50,000,000.

The interval of time during which the stimuli were presented to *S* was kept constant at two sec. by means of a thyratron timer circuit. The time interval was determined by varying the amount of charge on a condenser by means of a Helipot variable resistor.[1] Closing a master switch started the thyratron timer which in turn activated precision relays whose opening and closing time are less than one msec. The relays simultaneously started the chronoscope and presented the tone or light stimuli or both. When *S* responded by closing his telegraph key, a thyratron keying circuit closed another precision relay, which stopped the chronoscope oscillator. The stimulus, however, remained on for its pre-set two-sec. interval. The design of the keying circuit made it impossible for *S* to affect the count by making more than one response; only the first response was recorded by the counter. A pilot light on *E*'s control panel provided a visual indication of every closure of the response key by *S*. When *E* had recorded the count (reaction time) on the data sheet, opened the master switch, and reset the chronoscope counter to zero, the apparatus was ready for the next trial. The overall precision for measuring time is very high. With fixed settings of the electronic timer, the standard deviation of readings of the chronoscope is of the order of 0.4 msec. Since such readings involve errors due to variation of the timer, the relays and the chronoscope, the measures of latency (reaction time) are certainly accurate to one msec.

Ss for this experiment were recruited from classes in elementary psychology at the University of Wisconsin by the offer of an additional point toward their final grade in the course for each hour of service. Ninety *Ss*, 51 men and 39 women, were each assigned by a table of random numbers to one of the nine different groups, so that there were 10 in each group. The preliminary instructions to all *Ss* upon entering the experimental room were as follows: "Sit down at the table, please. I'll adjust my apparatus for the experiment and be back in a few minutes and give you instructions. Please stay seated and do not touch or disturb the apparatus on the table."

E then went to the control room and performed one of the three following actions:

A. If *S* was in the sensory pre-conditioning group, 10 trials of light and tone were presented. *Ss* given this pre-training treatment will be designated hereafter by the letter A.

B. If *S* was in the control group for sensory generalization, no stimulus was presented, but *E* waited 120 sec. (the time interval needed to present stimuli in the other groups). *Ss* given this pre-training treatment will be designated by the letter B.

C. If *S* was in the control group for non-contiguous presentation of light and tone, 10 trials of tone alone were presented. *Ss* given this pre-training treatment will be designated by the letter C.

E then returned to the experimental room and gave one of the three following sets of instructions:

I. If *S* was in the group duplicating the instructions used in the experiment which first demonstrated sensory pre-conditioning in human *Ss* (3), he was given the following instructions: "This is an experiment in reaction time, as you know. Let's adjust the chair to the table so that you are comfortably seated when your right arm is rested on the table with the two forefingers of your right hand on the telegraph key. Press the key a few times and notice that it moves through a very short distance and that it does not need to be pressed hard. Your task is to press the key as fast as possible when the light in the cabinet comes on. You need press it only once and may release it immediately. Be careful that you rest your fingers lightly on the key so that it is not closed when the next light comes. I am going to place these electrodes on your left wrist. The apparatus is adjusted so that you will receive a weak electric shock on your wrist if your reaction time is slow. If you give a quick reaction, you will receive no shock, so respond just as fast as you can. I'll answer any questions you may have about the experiment when it is over." *Ss* who received these instructions will be designated by the Roman numeral I.

II. If *S* was in the group to be given very specific instructions, the following was presented (the section of the instructions in italics is the part differing from instructions I): "This is an experiment in reaction time, as you know. Let's adjust the chair to the table so that you are com-

[1] This make of variable resistor enables fine adjustment by allowing 5400° of rotation—i.e., 15 turns of a dial to cover the complete range of resistance.

fortably seated when your right arm is rested on the table with the two fore-fingers of your right hand on the telegraph key. Press the key a few times and notice that it moves through a short distance and that it does not need to be pressed hard. Your task is to press the key as fast as possible when the light in the cabinet comes on. You need press it only once and may release it immediately. Be careful that you rest your fingers lightly on the key so that it is not closed when the next light comes. *During the experiment the stimulus will change to a tone. You are not to respond to the tone.* I am going to place these electrodes on your left wrist. The apparatus is adjusted so that you will receive a weak electric shock on your wrist if your reaction time is slow. If you give a quick reaction, you will receive no shock, so respond just as fast as you can. I'll answer any questions you may have about the experiment when it is over." *Ss* who received these instructions will be designated by the Roman numeral II.

III. If *S* was in the group to be given freedom in choosing to respond or not to respond to tone in the post-test, instructions were presented as follows (the section of the instructions in italics is the part differing from that of both instructions I and II): "This is an experiment in reaction time, as you know. Let's adjust the chair to the table so that you are comfortably seated when your right arm is rested on the table with the two fore-fingers of your right hand on the telegraph key. Press the key a few times and notice that it moves through a very short distance and that it does not need to be pressed hard. Your task is to press the key as fast as possible when the light in the cabinet comes on. You need press it only once and may release it immediately. Be careful that you rest your fingers lightly on the key so that it is not closed when the next light comes. *During the experiment the stimulus will change to a tone. Do not decide in advance what you are going to do to the tone; do what seems natural.* I am going to place these electrodes on your left wrist. The apparatus is adjusted so that you will receive a weak electric shock on your wrist if your reaction time is slow. If you give a quick reaction, you will receive no shock, so respond just as fast as you can. I'll answer any questions you may have about the experiment when it is over." *Ss* who received the above instructions will be designated by the Roman numeral III.

While the instructions were being given, *E* adjusted the electrodes on *S*'s wrist, and demonstrated key pressing if it was necessary. No questions were answered, and *E* left the experimental room as quickly as possible. He then went to the control room and all groups were presented with the following sequence of stimuli: 30 trials of light, then 10 trials of tone, and finally 10 trials of light and tone in combination. *S*'s reaction time was recorded for every trial to which a response occurred. No shock was ever given in the experiment. The interval between trials was varied at random from 10 to 25 sec., except that between the 30th trial of light and the 1st trial of tone the interval was always 10 sec., as was the interval between the 10th trial of tone and the 1st trial of tone and light.

At the close of the experiment, *S* was asked why he responded or failed to respond to the

TABLE I

SUMMARY SCHEMA OF EXPERIMENTAL PROCEDURE

T = tone; L = light

Instructions	Treatment		
	A	B	C
I	Pre-training—10 T + L Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—none Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—10 T Training—30 L Post-test—10 T Post-post-test—10 T + L
II	Pre-training—10 T + L Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—none Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—10 T Training—30 L Post-test—10 T Post-post-test—10 T + L
III	Pre-training—10 T + L Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—none Training—30 L Post-test—10 T Post-post-test—10 T + L	Pre-training—10 T Training—30 L Post-test—10 T Post-post-test—10 T + L

tone, and whether he expected to get a shock with tone. If he received treatment A, he was asked whether he saw the light and heard the tone before the experiment was started. If he received treatment C, he was asked whether he heard the tone. Answers to all questions were recorded on the data sheet.

One may treat the present experiment either as one of factorial design or as three parallel experiments on sensory pre-conditioning. Although the factorial design is not particularly meaningful for an analysis of results of sensory preconditioning, a 3×3 table of instructions versus pre-training treatment provides a clear summary of the experiment. Table I is a summary schema of the experiment treated from the standpoint of factorial design. Each cell of the table represents a group of 10 *Ss* for which there is a different combination of instructions and treatment. However, for analysis of results, the experiment is considered to be three parallel experiments on sensory pre-conditioning, for which the instructions in each case are different.

Quantitative Results

The experiment provides several measures by which the effect of instructions upon sensory pre-conditioning may be assessed. The frequency and latency of response to the 10 trials of tone in the post-test provide the primary data for analysis. The frequency and latency measures of response to the tone and light combined provide another set of data. For both the post-test and the post-post-test, relatively greater frequency or relatively shorter latency for groups IA, IIA, or IIIA, than for groups IB and IC, IIB and IIC, and IIIB and IIIC, respectively, would be evidence of sensory pre-conditioning. In the case of the post-test trials of tone, the assumption is made that if there is sensory conditioning of the tone and light during pre-training on treatment A, conditioning of the key pressing response to the light should result in elicitation of the same response by the tone. The assumption of sensory conditioning for *Ss* given treatment A is required for analysis of the post-post-test trials of light and tone, and an additional assumption must be made that sensory conditioning will result in summation, whereas with lack of sensory conditioning for pre-training treatments B and C, the instructions will provoke some conflict between light and tone and result in a decrement of response.

Table II presents the mean frequency of response and the standard error of the mean to the 10 trials of tone in the post-test for each of the nine groups of *Ss*. Since there are 10 *Ss* in each cell, the maximum mean frequency in any cell is 10. To check on the occurrence of sensory pre-conditioning as a function of instructions, the differences between group mean frequency of response were subjected to the *t*-test.[2] These results are given in Table III. The five percent confidence level was accepted as the criterion for statistical significance. Group IA gave a significantly greater frequency of response in the post-test than did either of its controls, Groups IB and IC. There is no reliable difference between these two latter groups. Group IIA gave a greater

TABLE II

Mean Frequency of Response to the 10 Trials of Tone in the Post-Test for Each Group

Instructions	Pre-Training Treatment					
	A		B		C	
	M	σ_M	M	σ_M	M	σ_M
I	4.3	1.05	0.8	1.05	1.2	1.05
II	1.4	1.05	0.0	1.05	0.4	1.05
III	6.2	1.05	6.3	1.05	6.2	1.05

[2] Both the standard error of the means and the standard error of the differences between means were computed from the error mean square obtained by analysis of variance and thus are the same for each mean and each difference. This procedure is described by Snedecor (6).

TABLE III

DIFFERENCES BETWEEN GROUP MEANS OF FREQUENCY IN RESPONSE TO THE 10 TRIALS OF TONE IN THE POST-TEST

Groups	IA–IB	IA–IC	IB–IC	IIA–IIB	IIA–IIC	IIB–IIC	IIIA–IIIB	IIIA–IIIC	IIIB–IIIC
$M_1 - M_2$	+3.5	+3.1	−0.4	+1.4	+1.0	−0.4	−0.1	0.0	+0.1
$\sigma_{Diff.}$	1.49	1.49	1.49	1.49	1.49	1.49	1.49	1.49	1.49
t	2.35	2.08	0.27	0.94	0.67	0.27	0.07	0.00	0.07
P	.02	.03	.39	.19	.25	.39	.47	.50	.47

frequency of response than its controls, Groups IIB and IIC, but neither difference is reliable at the five percent confidence level. For all practical purposes there are no differences between Groups IIIA, IIIB, and IIIC.

The mean reaction time measures of the responses to the post-test trials of tone were computed. The *t*-tests of the differences between means revealed no evidence of sensory pre-conditioning. Similar results were obtained for the frequency measure of the post-post-test trials of tone and light. All *Ss* responded to each of the 10 trials of the tone and light. This measure, then, provides no evidence of sensory pre-conditioning. Analysis of the differences in mean reaction time for the trials of the post-post-test shows it to be no index of sensory pre-conditioning.

Analysis of the answers by the *Ss* to the questions at the close of the experiment in general provides support to the results obtained in terms of the key pressing response. Since there was no appreciable difference in the results of the questions from the results for questions in the previous experiment (3), these data are not presented.

Since all *Ss* received 30 reaction-time trials to light, the experiment may also be considered in terms of these responses, and an analysis may be made of the effect of instructions and pre-training treatment. The last 10 trials of the training period were used for an analysis of variance (6). No significant variation was found for instructions, pre-training treatment, or the interaction of instructions and treatment. There is no evidence that reaction time to light is affected by either the instructions or pre-training treatments given in the experiment.

DISCUSSION

The only positive results of the experiment were in terms of the frequency of response to the post-test trials of tone, and sensory pre-conditioning was demonstrated only with the original instructions, the set designated by Roman numeral I in the present experiment. The results for instructions II are in the direction required to demonstrate sensory pre-conditioning, but the differences between Groups IIA and IIB, and Groups IIA and IIC are not statistically significant. The results for instructions III are completely negative.

Although the results of the experiment are meager in terms of instructions, the change in the location of the speakers, and the indefinite localization of the tone produced more efficient conditions for obtaining sensory pre-conditioning with instructions I. In the previous experiment (3) in which the same instructions were used, an *N* of 42 was required for each of the groups with different pre-training in order to obtain differences in frequency of response to the trials of the post-test which were reliable be-

tween the five and one percent confidence levels. In the present experiment, similar differences between the experimental and control groups were obtained for instructions I with a group *N* of 10. The mean frequency of response to the post-test trials of tone in the previous experiment was 3.0 and in the present experiment was 4.3. The difference between these two means is not statistically significant. In the previous experiment, 11 of the 42 *Ss* (26 percent) in the experimental group responded to 8 or more of the 10 trials of tone in the post-test, whereas in the present experiment 5 out of 10 (50 percent) in the experimental group gave a frequency of response of 8 or greater. The difference of 24 percent between these two proportions is statistically significant. These differences in favor of the present experiment show a clear gain in efficiency of producing sensory preconditioning. Although the gain has been attributed above to the change in locus of the loud speakers and the indefinite localization of the tone, it may be that the new electronic control apparatus also contributed to the gain. The electronic chronoscope could have had no effect, since the reaction time measures failed to provide a measure of sensory pre-conditioning.

The results of the present experiment are interpreted to favor a focus for further research on variables other than instructions. Further understanding of sensory pre-conditioning should be obtained by determining the effect of variables such as the intensity of the pre-training stimuli, the number of pre-training trials, the number of training trials and so forth.

Summary and Conclusions

1. To test possibilities of demonstrating sensory pre-conditioning in human *Ss* under more efficient conditions, three different sets of instructions were given. With each set, the occurrence or lack of occurrence of sensory pre-conditioning was determined by statistical comparisons of frequency and latency of response to a post-test of 10 trials of tone and a post-post-test of 10 trials of light and tone in combination between groups who received pre-training trials of (a) paired tone and light, (b) tone alone (control for non-contiguous presentation of tone and light), and (c) no pre-training trials (control for sensory generalization). The three sets of instructions used were: (1) a duplication of the original instructions wherein *S* was told to give a key-pressing response when a light stimulus occurred; (2) *S* was told to respond to light but not to tone; and (3) *S* was told the light stimulus would be followed by tone and to do what seemed natural when tone came.

2. More efficient control and measurement conditions were established by substitution of electronic control of stimulus presentation and duration, and electronic recording of reaction time for the previously used mechanical apparatus. Also, the tone producing speaker was removed from the visual field of *S* so as to isolate the tone and light stimuli and provide an indefinite localization of the source of the tone.

3. Reliable evidence for the occurrence of sensory pre-conditioning was demonstrated only under conditions wherein the original instructions were used. The other two sets of instructions were statistically non-significant. The reaction time measures gave negative results in all cases. Negative results were obtained for the frequency and reaction measures of the 10 post-post-test trials of light and tone.

4. Apparatus changes resulted in more efficient production of the sensory pre-conditioning effect. Comparison with the previous experiment (3) shows a smaller *N* required to give the effect. 42 *S*s per group were used before, 10 *S*s per group were used with the changed apparatus. *S*s giving eight or more responses to the post-test trials increased from 26 percent in the previous experiment to 50 percent in the present experiment.

5. The results are interpreted to favor the investigation of variables other than instructions for further understanding of the phenomena of sensory pre-conditioning.

6. The analysis made of reaction time latencies of the last 10 training trials to light showed no significant evidence that reaction time is affected by either the instructions or pre-training treatments given.

(Manuscript received May 7, 1948)

References

1. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, 25, 323–332.
2. Brogden, W. J. Tests of sensory pre-conditioning with human subjects. *J. exp. Psychol.*, 1942, 31, 505–517.
3. Brogden, W. J. Sensory pre-conditioning of human subjects. *J. exp. Psychol.*, 1947, 37, 527–539.
4. Karn, H. W. Sensory pre-conditioning and incidental learning in human subjects. *J. exp. Psychol.*, 1947, 37, 540–544.
5. Potter, J. T. A four tube counter decade. *Electronics*, 1944, 17, 110–113.
6. Snedecor, G. W. *Statistical methods.* (3rd ed.) Ames, Iowa, 1940.
7. Sterling, H. T. A simple "Wien bridge" audio oscillator. *QST*, 1946, 30, 29–32.

A FACTORIAL APPROACH TO THE ANALYSIS OF VARIANCES IN ESTHETIC JUDGMENTS

BY J. P. GUILFORD

University of Southern California

AND

J. W. HOLLEY

University of Illinois, Chicago Undergraduate Division

During the past 20 years there have been a number of investigations in which factorial methods have been applied to esthetic judgments made by groups of individuals, usually in an effort to discover fundamental variables in the determination of those judgments (2, 3, 4, 5). The results from rather isolated and diverse studies have been moderately promising but in most instances have not been followed up. A systematic set of esthetic variables and their relations to other psychological variables have consequently not been brought into the general body of psychology. Not many of the esthetic factors have been verified in two or more investigations, and in most instances the interpretation of such factors leaves much to be desired.

The study reported here does not attempt either to add to the list of factors or to achieve better interpretation of any already proposed. Its chief concern is with methodology, though the factors that came out as a by-product should be of some interest. The major aim of the investigation was to assume that certain hypothetical esthetic factors exist and that they do tend to determine human judgments of esthetic objects, then to attempt to increase the contribution of each hypothetical factor to judgments by giving the subjects special instructions emphasizing in turn each factor.[1]

In addition to our interest in the validity of the factorial approach to a study of the determiners of esthetic judgments, we regarded a related problem as being important. This is the problem of the objectivity versus subjectivity of judgments. This problem was raised most clearly by Adams (1) who also proposed an objectivity-subjectivity ratio, an index that could be applied to any set of judgments. Adams's ratio is based upon indices (correlation coefficients) of self-consistency (reliability coefficients) and of 'group consistency' (intercorrelations between judges). The objectivity of any particular set of judgments of objects of a class may be defined (to satisfy Adams's conception) as the extent to which those

[1] This report is based upon two master's theses done under the direction of the senior author: Anita F. Klauss, *A factorial analysis of some aesthetic judgments*, 1938 (on file in the University of Nebraska library) and J. W. Holley, *A factorial study of some aesthetic judgments*, 1947 (on file in the University of Southern California library). Miss Klauss designed the original study and compiled the data, including some of the intercorrelations upon which this report is based. Mr. Holley computed many additional intercorrelations and carried out the two factor analyses reported here. We are indebted to Miss Klauss for the use of her data and hereby express our appreciation. The term 'esthetic' is used in a broad, psychological sense in this report.

judgments are determined by the objects. Subjectivity is correspondingly defined as the extent to which the judgments are unique to particular judges. As 'group consistency' approaches self consistency, the ratio approaches +1.00, which indicates complete objectivity.

In our opinion, Adams's index of objectivity is unsatisfactory for several reasons. There are questions of the mathematical propriety of using ratios of correlation coefficients without giving adequate underlying rationale for doing so. There are also questions concerning the apparent oversight of error variances in judge's ratings and their elimination from consideration in evaluating objectivity. It is not worth while to belabor these objections. We will, instead, propose a substitute conception based upon factor theory.

Assume that we have the esthetic judgments of a sample of N objects made by n judges on some metric scale. A tabulation of these metric values, one column for each judge and one row for each object, may be regarded as a matrix of order $N \times n$. Such a matrix might be the starting point for either an analysis-of-variance or a factor-analysis approach. To provide a more complete picture, however, both approaches would require at least one additional matrix of similar judgments of the same objects by the same judges.

Let us take the factor-analysis approach. Analysis of variance here would tell us whether certain contributions to total variance are statistically significant but would apportion contributing variances among *experimental* variables. If we wish to refer component variances to *underlying* variables (factors) which our experimental controls fail to segregate, we must resort to factor analysis. We could analyze the matrix of judgments, but the more usual procedure is first to compute intercorrelations. In this step we would have a choice between correlating rows (objects) or columns (judges). The latter is called 'inverted' factor analysis or the 'Q-technique' (7). It is the one more commonly applied in the analysis of esthetic judgments. One reason may well be that there are more objects than judges, in which case in using the Q-technique the size of sample is larger and the number of variables to be analyzed is smaller, both being important operational considerations. The factors so determined are more directly assignable as properties of the judges but can also be related to groups of objects, as will be shown later.

Having solved a particular factor-analysis problem by the Q-technique, we have as the result a matrix of common-factor loadings, with n rows (one for each judge) and r columns (one for each common factor). The square of each common-factor loading indicates the proportion of the total variance in any particular judge's ratings of the set of objects that can be attributed to a single factorial source. The sum of each judge's squared factor loadings is known as communality. It is the sum of his common-factor variances. We propose that this statistic be used as the index of each judge's degree of objectivity in making his set of judgments of the particular set of objects. Since each person's communality is derived from his correlations with all other judges in the experiment, it indicates the extent to which his evaluations were similarly determined to those of the other individuals. It is a shorthand expression of his degree of agreement with the group of judges. Objectivity becomes defined opera-

tionally as the extent of inter-person agreement as indicated by the sum-total of common-factor variance. In order to obtain a single index of objectivity for a set of objects, an index that is representative over all persons in the experiment, an arithmetic mean of the communalities might be used.

The remaining proportion of the variance (total variance minus communality) in each individual's judgments may be regarded as being unique to him, but not all of it should be legitimately regarded as subjective. It is further analyzable into *specific* variance and *error* variance. The specific variance is indicated by the difference between an estimate of reliability for a judge and his communality. Reliability measures the proportion of total variance that is non-error. Some of the non-error variance is shared with other judges (communality) and some is unique to the individual judge (specific variance). The error variance is no genuine or enduring property of the judge. It is indicated by the difference between the reliability coefficient and +1.00 (the total variance). In a case such as esthetic judgments, the reliability for a single judge is ordinarily of the 'test-retest' type; a correlation between two experimentally 'independent' series of ratings of the same set of objects. In results to be reported later in this article, examples of the apportionment of each judge's variances will be illustrated.

The results to be presented also have some bearing upon the relative effectiveness of experimental instructions in determining judgments when objects are complex. Some judgments were made under the instructions to let certain aspects of the objects—color, design, or theme—determine the evaluation of the object. It will be shown how apparently poorly the instructions were followed, either by failure of the judges to abstract the aspect intended, by the fact that these three aspects do not represent fundamental dimensions of esthetic response, or by other reasons which we do not know.

Plan of the Investigation

The objects.—The material chosen for use in this study was in the form of designs on the backs of playing cards. For better experimental control, especially designed artistic objects could have been used. Perhaps the chief virtues of playing cards are their ready availability and the ease with which they can be manipulated by an observer. There are other advantages in their use. They represent considerable variety of color combinations, of designs, and of subject matter. Furthermore, they come in rather comparable pairs so that duplicate studies can be made. Although the designers of playing-card backs attempt to produce popularly pleasing results, there is probably a considerable range of artistic merit represented, particularly when several aspects of artistic merit may have a bearing upon the reactions of different individuals. The use of artistic objects that were prepared for commercial objectives also carries with it the possibility of showing the applicability of the factorial approach to problems of marketing, though this was incidental to the present investigation. The two sets of 115 cards each will be known hereafter as packs A and B.

The subjects.—The judges were 12 students, three from each of four colleges at the University of Nebraska. None were students of fine arts. Six were males and six were females. In the symbolism following, the women are odd-numbered and the men are even-numbered.

The rating method.—The esthetic judgments were made by the application of the method of equal appearing intervals with seven categories. In making any particular set of judgments for the 115 designs, the observer began by looking through the entire set to find the card design he liked best of all and the design he disliked most (or liked least). These he placed at the ends of his row of seven piles as anchoring points. As the observer decided where to place each card, he placed it face down in its pile.

The instructions.—There were four different instructions used by different combinations of observers and packs (A or B). These instructions are briefly defined as follows:

I. General impression: "Judge on the basis of *general impression*, the *total effect* of the card upon you."

II. Color effect: Same wording as in I, except for the term 'color effect' in place of 'general impression.'

III. Artistic pattern or design: Also called 'form or decorative pattern.'

IV. Subject matter, theme, meaning, or topic.

The reasons underlying the framing of the four instructions are as follows: It was hypothesized that the most obvious aspects of the objects—color, design, and theme—would make themselves evident in some way in the factor analysis. Perhaps each one would provide a separate factor or a set of factors. It was thought that if all judgments were obtained under the general-impression instruction alone, the identification of factors might be more difficult. If factors tended naturally to separate along the lines of the special instructions it was believed that the interpretation of factors would be facilitated.

The design of the study.—It was felt to be important that the reliability of judgments made by each subject for each set of cards under each instruction be estimated. This called for a repetition of each set of judgments under each condition. In order that each subject might not repeat judgments of the same set of cards even under different instructions too many times (and thus develop stereotyped reactions and perhaps some degree of satiation for the task and for particular designs), he judged each set under only two instructions. In order to prevent memory factors from unduly influencing the ratings, each subject rated each set of designs only once on a single day. Table I shows the first two days' schedule. The third and fourth days' schedule was the same in completely reversed, counterbalanced order.

TABLE I

ORDER OF PRESENTATION OF COMBINATIONS OF INSTRUCTIONS AND SETS OF DESIGNS FOR THE 12 SUBJECTS *

Subject	First Day		Second Day	
1	IIA	IIIB	IVB	IA
2	IIIA	IIB	IVB	IA
3	IVA	IIB	IIIB	IA
4	IB	IIIA	IIA	IVB
5	IB	IVA	IIA	IIIB
6	IB	IVA	IIIA	IIB
7	IVB	IA	IIA	IIIB
8	IVB	IA	IIIA	IIB
9	IIIB	IA	IVA	IIB
10	IIA	IVB	IB	IIIA
11	IIA	IIIB	IB	IVA
12	IIIA	IIB	IB	IVA

* Roman numerals refer to instructions, letters to sets of designs.

Statistical operations.—The reliability for each set of conditions (subject-instruction-deck) was estimated by means of a product-moment correlation with subsequent application of the Spearman-Brown formula. For the intercorrelation data, the two sets of ratings obtained under each set of conditions were summed by pairs. In this report, only data derived from deck A will be considered. The factor analyses were carried out by Thurstone's centroid method followed by rotation of reference axes so as to minimize the number of negative loadings and maximize the number of zero loadings. The rotations were made according to Zimmerman's graphic method (8). Two separate analyses were carried out, each involving 12 variables. There were actually 24 variables based upon deck A, but they were subdivided for a very important reason. Examination of the entire 24×24 correlation matrix showed that the pairs of variables each involving the same judge intercorrelated conspicuously higher than pairs derived from different judges. This would insure quite a number of 'doublet' factors, each restricted to one judge. By subdividing the total matrix so that the same individual was not represented more than once in the same analysis, we avoided these doublets. The same 12 judges are represented once each in the two correlation matrices analyzed so that comparisons can be made in the two factor configurations. We also present a special consideration of the intercorrelations for the same judges under two instructions by correcting them for attenuation. It is of interest to see whether such corrected coefficients are close to +1.00. Wherever they are, the conclusion would seem to be that the individual in question is either unaffected by instructions, or (if one of his instructions is general) he adopted for himself under general instructions a set similar to that induced by one of the special instructions.

These were the general statistical operations. In connection with the extraction of centroid factors, a new procedure for assuring the extraction of maximal variance each time was devised and used by the junior author, who describes the procedure elsewhere (6). He also applied a new method for estimating new guessed communalities after extraction of a first set of centroid factors. This method is also described elsewhere.[2] Later in this article we will describe a procedure that was used as an aid to the identification of the factors and which may prove to have more general application. This was also developed by the junior author.

[2] See reference to Holley's thesis in footnote 1.

Results

Reliability of the judgments.—Table II presents the estimates of reliability of the judgments in the upper right half of Sub-table B. The individuals and the instructions to which these reliability coefficients belong are shown in the corresponding cells in the upper right part of Sub-table A. The cells should be interpreted as follows: Individuals 1 and 7 had had both instruction I and instruction II. For individual 1, the reliabilities under instructions I and II were .86 and .82, respectively; those for individual 7 were .81 and .86, respectively. The reason for this unusual arrangement of reliability data is that it provides a more ready comparison of intercorrelations between instructions for the same individual. These intercorrelations appear in the lower left halves of the tables, uncorrected in sub-table A and corrected for attenuation in sub-table B. Concerning the reliabilities it may be said that most of them are relatively high, indicating sufficient 'true' variance in the ratings to justify their use in a factor analysis.

Within-individuals correlations.—We will refer to the correlation between two sets of judgments made by the same person under two different instructions as 'within-individual' correlations. With one exception, the correlations within individuals are rather substantial or high. When corrected for attenuation, six of the 12 are close to +1.00, in which cases the interpretation is that variation of instructions had little effect upon the judgments. The fact that four of the six coefficients are greater than +100 may mean that the reliability coefficients were underestimated (probably due to errors of coarse grouping) or it may be due to errors of sampling in either reliability estimates or intercorrelations. It is rather certain that the other six individuals were affected to some extent by variations of instructions, one individual (number 6) particularly.

The factorial results.—The two correlation matrices are represented in Tables III and IV. The centroid loadings, with six factors extracted, appear in Tables V and VI. Extractions were continued through six factors but only the first five were rotated because the sixth seemed to offer no improvements in the adjustments to a configuration approaching simple structure. The extractions were repeated a second time with empirical estimates of communalities so as to reduce the discrepancies be-

TABLE II

Intercorrelations Among Pairs of Sets of Judgments by the Same Individual; Estimates of Reliability of Repeated Judgments; and Intercorrelations Corrected for Attenuation *

Sub-table A

Instruction:	I		II		III		IV	
I			1	7	2	8	3	9
II	.57	.74			4	10	5	11
III	.59	.80	.78	.78			6	12
IV	.95	.87	.76	.57	.26	.85		

Sub-table B

Instruction:	I		II		III		IV	
I			.86	.81	.86	.84	.93	.75
			.82	.86	.59	.82	.85	.90
II	.68	.89			.78	.73	.91	.69
					.82	.79	.94	.85
III	.83	.96	.98	1.03			.77	.77
							.89	.88
IV	1.07	1.06	.82	.74	.31	1.03		

* Key to the table: Sub-table A, lower left, contains the uncorrected intercorrelations; upper right contains the numbers of the corresponding individuals for whom the intercorrelations stand. Sub-table B, lower left, contains the corrected intercorrelations; upper right contains the reliability estimates.

tween estimated values and computed values. The rotated factor loadings with communalities are given in Tables VII and VIII.

It is of interest, first, to see whether the two factor matrices (rotated) present similar configurations so that factors in the one may be identified

TABLE III

First Correlation Sub-Matrix

Condition	1AI	2AI	3AI	4AII	5AII	6AIII	7AII	8AIII	9AIV	10AIII	11AIV	12AIV
1AI	—											
2AI	−.09	—										
3AI	.27	.22	—									
4AII	.20	.40	.32	—								
5AII	.30	−.26	.08	.07	—							
6AIII	.00	.32	.08	.39	.01	—						
7AII	.04	.10	.09	.35	.52	.37	—					
8AIII	−.09	.52	.01	.20	−.32	−.03	−.13	—				
9AIV	.11	−.51	−.07	.00	.60	.05	.36	−.28	—			
10AIII	−.09	.00	−.07	.09	.38	.15	.42	−.12	.34	—		
11AIV	.22	.15	.11	.36	.10	.34	.21	.07	.18	.23	—	
12AIV	.20	.45	.27	.51	.29	.40	.20	.16	.06	.20	.08	—

TABLE IV

Second Correlation Sub-Matrix

Condition	1AII	2AII	3AIV	4AIII	5AIV	6AIV	7AI	8AI	9AI	10AII	11AII	12AIII
1AII	—											
2AIII	−.05	—										
3AIV	.17	.24	—									
4AIII	.10	.20	.33	—								
5AIV	.09	−.18	−.06	.00	—							
6AIV	.05	.11	−.04	−.02	.13	—						
7AI	−.02	.32	.03	.25	.57	.05	—					
8AI	−.04	.34	.00	.29	−.51	.08	−.19	—				
9AI	.09	−.15	−.10	.09	.66	.32	.34	−.27	—			
10AII	−.18	.31	.03	.09	.46	−.11	.50	−.18	.27	—		
11AII	.40	.08	−.03	.34	−.06	.06	.12	.17	.11	−.06	—	
12AIII	.18	.41	.26	.31	.09	.38	.32	.08	.02	.13	.21	—

TABLE V

Centroid Factor Loadings for the First Sub-Matrix

Condition	K_1	K_2	K_3	K_4	K_5	K_6
1AI	.324	−.108	.566	.159	−.059	−.046
2AI	.229	−.745	−.482	.225	.190	−2.43
3AI	.295	−.271	.184	.154	−.163	−.033
4AII	.589	−.422	−.055	−.063	−.127	.092
5AII	.633	.592	.175	.346	.081	−.019
6AIII	.508	−.212	−.252	−.303	−.205	−.062
7AII	.652	.277	−.252	−.089	−.023	−.148
8AIII	−.065	−.460	−.205	.068	.271	.130
9AIV	.372	.582	.132	−.108	.019	.191
10AIII	.422	.344	−.259	−.084	.213	.066
11AIV	.492	−.180	.208	−.482	.393	−.069
12AIV	.609	−.317	−.168	.314	−.052	.289

with factors in the other. The systematic imposition of special instructions upon the judges is the same for both matrices and the same 12 individuals are represented in both. Furthermore, the high intercorrelations as seen in Table II would lead us to expect similar configurations in

TABLE VI

CENTROID FACTOR LOADINGS FOR THE SECOND SUB-MATRIX

Condition	K_1	K_2	K_3	K_4	K_5	K_6
1AII	.287	−.301	−.485	−.337	.025	.207
2AIII	.343	−.387	.559	.308	.117	.147
3AIV	.246	−.309	.198	−.334	−.258	−.118
4AIII	.377	−.344	.182	−.218	.154	−.297
5AIV	.651	.687	−.171	−.029	−.062	−.075
6AIV	.271	−.155	−.287	.537	−.192	−.156
7AI	.676	.246	.249	.029	.127	.083
8AI	−.153	−.569	.162	.173	.303	−.207
9AI	.516	.381	−.350	.114	.085	−.276
10AII	.448	.384	.433	.060	.152	.069
11AII	.292	−.342	−.242	−.112	.444	.102
12AIII	.593	−.428	.075	.154	−.290	.119

TABLE VII

ROTATED FACTOR LOADINGS AND COMMUNALITIES FOR THE FIRST SUB-MATRIX

Condition	K_1	K_2	K_3	K_4	K_5	h^2
1AI	−.019	−.017	.680	.111	−.050	.477
2AI	.807	−.421	−.100	−.002	.294	.924
3AI	.256	−.127	.374	.135	−.106	.251
4AII	.538	−.084	.250	.446	.039	.560
5AII	.156	.814	.473	−.097	−.074	.925
6AIII	.413	.062	−.066	.572	−.029	.507
7AII	.371	.590	.016	.320	.006	.588
8AIII	.292	−.347	−.101	−.098	.330	.335
9IV	−.153	.665	.160	.162	−.041	.519
10AIII	.202	.561	−.093	.123	.194	.418
11AIV	.036	.110	.252	.548	.574	.706
12AIV	.714	.051	.288	.100	−.004	.606

TABLE VIII

ROTATED FACTOR LOADINGS AND COMMUNALITIES FOR THE SECOND SUB-MATRIX

Condition	K_1'	K_2'	K_3'	K_4'	K_5'	h^2
1AII	−.058	.128	.514	−.270	.398	.514
2AIII	.470	−.180	.004	.635	.144	.677
3AIV	.022	−.090	.547	.238	−.076	.370
4AIII	.040	−.020	.352	.371	.301	.355
5AIV	−.024	.960	.008	.058	−.103	.938
6AIV	.626	.160	−.042	−.250	.091	.491
7AI	.132	.560	.050	.496	.088	.588
8AI	.188	−.530	−.062	.161	.376	.487
9AI	.103	.705	−.031	−.112	.170	.551
10AII	.010	.444	−.163	.569	−.079	.554
11AII	−.007	.047	.172	.033	.660	.469
12AIII	.588	.103	.473	.219	.088	.638

the two analyses. It was possible to identify four parallel factors rather clearly in the two matrices, and paired factors have been given identical numbers. Only factor four is not similar for the two matrices. Only factor two proves to be a clearly bipolar factor, the others having no strong negative loadings.

It is of interest to note whether the factors represent a clustering of variables by instructions, particularly along the lines of the three special instructions—for color, design, and theme. Inspection of the rotated loadings shows definitely that there was no such clustering. This is additional evidence that the special instructions were generally ineffective in producing factors. Factors undoubtedly exist in the realm of these judgments but they do not follow along the lines laid down by the instructions. As exhibited by the judgments of these 12 individuals, there seem to be no unitary variables identifiable as determination of preferences by color versus design versus theme.

This does not preclude *particular* color factors or design factors or theme factors. If there were more than one of each of these types operating, the clustering of the type anticipated would probably not occur.

The clustering is definitely more by individuals. Inspection of Tables VII and VIII in connection with Table II will show that those persons who have the highest within-individuals intercorrelations (subjects 3, 4, 8, 9, 10, and 12) also exhibit the greatest degree of similarity of factor loadings in Tables VII and VIII.

There is only a slight suggestion of sex differences in this clustering, in that males are conspicuous among those having strong positive loadings in factor two. Each variable represents not purely an individual but an individual combined with a certain instruction. But since we have shown instruction to be at best a very weak determiner of intercorrelations, the nature of the factors will have to be sought in the communality between individuals and in how individuals similarly evaluated certain types of designs.

Interpretation of the factors.—In order to determine the nature of the factors, a system of weighted judgments was employed. The product of the factor loading and the rating given by the individual to a particular card was obtained. The rating of the card for a particular factor was then regarded as the sum of the products for that card for the 12 individuals. The procedure is only roughly indicative, but it serves for the purposes of 'scoring' the cards for the factors. An examination of the cards with high and low 'scores' in the different factors should give some clues to the nature of the factors.

The cards were placed in rank order for each factor, laid out in a row and inspected. From this inspection the following suggestions can be made:

Factor I: This seems to be a special theme factor and it has been tentatively named 'romantic-adventure.' The picture having the highest score is an old sailing ship. The second is a picture of an orchid. The next three in the list present pictures of young couples in romantic situations. In some of the high-scoring cards there is a suggestion of 'escape from reality' into a setting of romantic adventure. Among the highest ten cards, six pictures are of couples.

Factor II: This seems to be a special design factor. For both submatrices, the highest-scoring cards are representative of simple, modern de-

sign. In this type of art, details are kept at a minimum and the figures are conventional rather than realistic. The bipolarity of this factor can be interpreted by supposing that individuals with negative loadings in this factor tend to prefer more complicated or realistic designs.

Factor III: In both sub-matrices, this factor seemed to be a 'feminine taste' or 'feminine interest' variable. The highest-scoring cards suggest social situations such as teas, parties, and dances. Delicacy of line and coloring seemed also to be a prominent feature. Women tended to have higher loadings in this factor than did men.

Factor IV: In both sub-matrices the high-scoring cards seem to indicate wealth, luxury, play, and recreation.

Factor V: This factor seems to reflect love of the outdoors, both in its vigorous and its restful aspects.

The general picture presented by these factors is one of relatively strong determination of judgments by motivational aspects as expressed in particular themes, with the one exception in which the factor seems to be a genuinely esthetic one of preference for simple and somewhat conventional or stylized design. There is a suggestion in this general result that judgments of preference for pictures such as playing-card designs might be used as a kind of projection test. With the items scored for their 'loading' with certain motivational factors, the individual's judgments might well be indicative of his position with respect to those factors. The scoring of such a test would be based upon empirical data and hence be very much more objective than that of present forms of projective tests. There is also the suggestion that judgments of pictures can be indirectly indicative of basic interests underlying vocational or avocational satisfactions. An interest test of this type would avoid some of the objections to present verbal tests.

The objectivity of judgments.—At the beginning of this report it was proposed that the degree of objectivity of an individual's judgments be indicated by his communality. His communality estimates the proportion of the total variance in his judgments that is related to variance in the judgments of other individuals. When the communality is broken down into components, each component being contributed by or attributable to a certain common factor, we have a further refinement of definition of any person's objectivity.

The list of reliabilities, communalities, unique, specific, and 'error' variances are given in Tables IX and X for the two sub-matrices. As usually defined in factor theory, the unique variance is the proportion of the total variance *not* attributable to common factors. Specific variance indicates self-consistent performance of the individual that he does not share with other individuals. In this kind of situation, 'error' variance might better be called 'inconsistency' variance. It

TABLE IX

RELIABILITIES, COMMUNALITIES, UNIQUENESSES, SPECIFICITIES, AND ERROR VARIANCES FOR THE FIRST SUB-MATRIX

Condition	r_{aa}	h^2	K_u^2	K_s^2	K_e^2
1AI	.860	.465	.535	.395	.140
2AI	.860	.926	.074	−.066	.140
3AI	.930	.245	.755	.685	.070
4AII	.780	.548	.452	.232	.220
5AII	.910	.909	.091	.001	.090
6AIII	.770	.501	.499	.269	.230
7AII	.860	.575	.425	.285	.140
8AIII	.820	.336	.664	.484	.180
9AIV	.900	.506	.494	.394	.100
10AIII	.790	.415	.585	.375	.210
11AIV	.850	.703	.297	.147	.150
12AIV	.880	.601	.399	.279	.120

TABLE X

RELIABILITIES, COMMUNALITIES, UNIQUENESSES, SPECIFICITIES, AND ERROR VARIANCES FOR THE SECOND SUB-MATRIX

Condition	r_{aa}	h^2	K_a^2	K_s^2	K_e^2
1AII	.820	.523	.477	.297	.180
2AIII	.590	.689	.311	−.099	.410
3AIV	.850	.374	.626	.476	.150
4AIII	.820	.365	.635	.455	.180
5AIV	.940	.930	.070	.010	.060
6AIV	.890	.504	.496	.386	.110
7AI	.810	.597	.403	.213	.190
8AI	.840	.495	.505	.345	.160
9AI	.750	.554	.446	.196	.250
10AII	.730	.562	.438	.168	.270
11AII	.690	.471	.529	.219	.310
12AIII	.770	.649	.351	.121	.230

should be remembered in this particular study that each variable is not merely an individual judge, but a judge combined with a certain instruction. Since we have concluded that instructions as given were rather impotent, the communalities that exist are probably mostly attributable to the properties of the individuals and their self-imposed sets. This is true in the case of at least six judges. On this basis we may say that individuals 2 and 5 show exceptional agreement with other judges in the first sub-matrix, and individual 5 also in the second one. We may also say that individuals 3 and 8 show exceptionally low agreement with other judges in the first sub-matrix, and individuals 3 and 4 in the second. Other individuals exhibit moderate degrees of agreement with their fellows in this study in both matrices. The consistency of communality as between matrices may be described as fair.

There would necessarily be some cerrelation between communality and reliability of judgments, since high communality cannot exist without high reliability. Low communality, however, can exist in spite of high reliability. The latter situation is shown by individuals 3, 4, and 8 (3 in both matrices). Such a gap between reliability and communality indicates considerable determination (of judgments) that is peculiar to the individual. Had other individuals been added to the list, new common factors might have been brought to light. This suggests that the number of factors that might be isolated and that might serve as descriptive of persons as expressed in esthetic judgments may be rather large. As for the 'error' variance, we are interested in it only as a complement to the self-consistency (reliability) variance.

In Tables IX and X, there is no segregation of the common variance among the common factors. Such a segregation would entail merely the squares of the factor loadings given in Tables VII and VIII. As an example of the total variance of judgments of individual 2 under instruction I, a proportion equal to .65 is attributable to factor one (romantic-adventure interest), .18 to factor two (liking for *complex* designs, because his loading is negative), .01 to factor three (masculine interest, since the loading is negative), .00 to factor four (interest in luxury), and .09 to factor V (outdoor life). The common-factor variances for this individual-instruction variable exhausts all except the 'error' variance, which amounts to .07 of the total. Individual 2 was completely objective in his judgments. These statements presuppose correct solution of the factor problem and correct interpretation of the factors. It is merely illustrative.

It is suggested that such descriptions of variables might well be utilized more often that they are because they are more directly interpretable as additive amounts of contribution of each factor to each experimental variable. We do not take the

space to present complete tables of proportions of variances in each variable attributable to each common factor here because there is little interest in the factors and the variables as such. There are other experimental situations in which such a complete listing of variances would be quite meaningful and useful, as, for example, in describing tests.

Practical implications.—The present study suggests that the factorial approach might be used to analyze the public demand for articles that appeal for various distinguishable reasons. For example, had we started out with the problem of why different kinds of playing-card designs are popular and with what types of customers, with a few changes in experimental design the answers could probably have been found. By uncovering the variables that enter into popular preferences for a kind of article, and by determining their weights in demands of customers, we could make suggestions to designers and to venders of playing cards as to what aspects to stress in designing and in buying. Much of the trial and error now involved in attempting to fathom the public appeal of many commodities—clothing, housing, and entertainment, to mention a few—could probably be avoided.

Summary and Conclusions

This study was undertaken primarily with the motive of throwing light upon the validity of the factor-analysis approach to the analysis of the variables underlying esthetic preferences. Efforts were made to bring out presupposed factors by imposing specific instructions upon the 12 judges who made repeated ratings of 115 playing-card designs.

The anticipated factors—common likings for colors, for designs, and for themes—did not emerge. Instead, five factors were found which could be interpreted as four theme factors and one special design variable. Special instructions in general proved to be rather ineffective; individuals gave highly similar ratings when instructions were changed.

It was proposed that each individual's common-factor variance be used as an index of his degree of objectivity (agreement with other judges) and that his specific variance be used to indicate his subjectivity. It was also proposed that judgments of esthetic objects, when scored for empirically determined factors, which in the present study seem to be mostly motivational, would serve as an excellent basis for more objective projection tests and tests of interests. It is also suggested that the factorial approach may have important contributions to make in the field of marketing research.

(Manuscript received March 26, 1948)

References

1. Adams, H. F. An objectivity-subjectivity ratio for scales of measurement. *J. soc. Psychol.*, 1930, 1, 122–130.
2. Beebe-Center, J. G. General affective value. *Psychol. Rev.*, 1929, 36, 472–480.
3. Eysenck, H. J. Type factors in aesthetics. *Brit. J. Psychol.*, 1941, 31, 262–270.
4. Eysenck, M. D. An experimental and statistical study of olfactory preferences. *J. exp. Psychol.*, 1944, 31, 246–252.
5. Harsh, C. M., Beebe-Center, J. G., & Beebe-Center, R. Further evidence regarding preferential judgments of polygonal forms. *J. Psychol.*, 1939, 7, 343–350.
6. Holley, J. W. A note on the reflection of residuals in the centroid method of factor analysis. *Psychometrika*, 1947, 12, 263–266.
7. Stephenson, W. The inverted factor technique. *Brit. J. Psychol.*, 1935–36, 26, 344–361.
8. Zimmerman, W. S. A simple method for orthogonal rotation of axes. *Psychometrika*, 1946, 11, 51–55.

THE EFFECTS OF REPEATED ROTARY ACCELERATION ON THE OCULO-GYRAL ILLUSION *

BY BRANT CLARK

San Jose State College

AND

KENNETH MacCORQUODALE

University of Minnesota

INTRODUCTION

In the last quarter of a century many studies have been made which have demonstrated the habituation of nystagmus to repeated rotary acceleration. The initial impetus to investigate this problem arose in connection with investigations of the reliability and general usefulness of the nystagmus test as a selection device for aviators. The first systematic study of the habituation of nystagmus to rotary acceleration was made by Dunlap and Bentley (5, 22) and this was shortly followed by similar studies by Griffith (15-18) and Holsopple (21–23). More recent studies have also reported habituation of nystagmus following repeated acceleration, notably those of Mowrer (32–37) and of Fearing and his associates (8, 9, 41). A few of these studies have used human subjects (3, 16, 21, 22, 27, 38) while others have been concerned with pigeons (9, 33–37, 41), rats (4, 9, 15) and rabbits (30, 36, 40). Most of the investigators have used rotary acceleration as the habituating stimulus, but a few have studied the effects of caloric stimulation (6, 10, 27). These studies are in almost unanimous agreement that nystagmus following angular deceleration is reduced to a marked degree by repeated rotary acceleration and deceleration and that in some cases it may even be reduced to zero. Habituation of past]pointing has also been reported (16, 34). The early observers who found no habituation were attempting to obtain habituation from rotation per se rather than a change in rate of rotation (i.e., angular acceleration) which is the adequate stimulus to the semicircular canals. Since the semicircular canals are generally conceived as mediating the nystagmus response, the disappearance of nystagmus during prolonged rotation at a constant rate involves neither habituation nor adaptation in any true sense; it depends on the dissappearance of the adequate stimulus rather than an adjustment to the stimulus. However, two careful studies of humans have failed to demonstrate marked habituation of nystagmus. One of these by Mowrer (38) examined angular acceleration as the stimulus and another (27) used caloric stimulation. Still another study by Dorcus also failed to demonstrate habituation in rats (4), and Prince obtained negative results in cats (35). The experimental evidence is, nevertheless, overwhelming in support of some reduction of nystagmus following repeated angular accelerations and decelerations. This applies particularly to animals having head nystagmus.

Habituation has also been reported very generally for certain subjective effects of angular acceleration. Numerous investigators report that dizziness, sensations of falling, nausea, vertigo and other organic effects tend to decrease or actually disappear following successive exposures to angular acceleration (5, 15, 17–19, 24). Griffith (16) and Mowrer (34) present evidence that there is also a reduction of 'visual vertigo.' However, several studies have also demonstrated that the *equilibrium* of subjects, both

* The experimental work for this study was done at the U. S. Naval School of Aviation Medicine and Research, Pensacola, Florida. The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the view of the Navy Department or the Naval Service at large.

human (22) and animal (34, 41) is not modified by repeated rotary acceleration, nor is the illusion of turning (21, 28, 29, 39).

The possibility of habituation to repeated angular acceleration has become a matter of concern to the writers and to other workers at the Naval School of Aviation Medicine and Research, since there has issued from this laboratory a series of papers (1, 2, 12, 13, 28, 29) on the effects of these accelerations on visual and non-visual space perception. Most of these studies utilized the same observers who were rotated in Barany chairs, Link Trainers, a human centrifuge, and airplanes, and therefore certainly had the opportunity to become habituated, if anyone ever did. Nevertheless, in a series of observations covering more than a year, no appreciable decrement appeared in the reports of these subjects. Moreover, several of these studies were specifically concerned with the oculogyral illusion,[1] which has been hypothesized as a perceptual result of nystagmus, an hypothesis which must be modified if nystagmus habituates to zero. However, these studies, which have seen in the oculo-gyral illusion a threat to pilot orientation, have involved angular accelerations and decelerations which were in most cases small in magnitude and were distributed over an extended period in time. Informal studies of a number of Naval aviators also showed that they observed the oculo-gyral illusion. It was believed, therefore, that in view of the preponderance of evidence for habituation of nystagmus and certain other effects normally associated with rotatory acceleration, it would be profitable to make a systematic study of the effects of repeated rotary accelerations on the occurrence of the oculo-gyral illusion.

[1] The apparent motion of a fixed target in the dark following angular acceleration or deceleration. The effect is similar to what earlier writers termed 'visual vertigo.'

Apparatus and Procedure

All of the observations were made in an air-conditioned, dark room where each subject was exposed to 276 accelerations, 276 decelerations and 5017 complete 360° turns during the experiment. The subjects were rotated in a Link Trainer, which was modified so that it could rotate only about a vertical axis. Communication between the experimenter and the observer was possible by means of the standard Link Trainer intercommunication system. The Link Trainer was controlled by the experimenter in an adjacent room and was rotated to the subjects' right at either three rpm or 20 rpm. These two rates of rotation were used because they produced respectively a relatively weak and a relatively strong accelerative force. These would serve to test the hypothesis that habituation could occur in different degrees for different accelerative forces.

The box which served as a fixation object was $8 \times 8 \times 14.5$ cm. and contained a light source. The box was opaque except where pierced along its edges by three-mm. holes. The holes were one cm. apart and the box was tilted so that the observer had a tri-dimensional view of the stimulus. The target box was mounted on a platform which turned with the Link Trainer and was two meters from the subject's eyes. This particular three dimensional target was selected because a previous study had shown that it reduced autokinesis to a minimum (12).

The adequate description of accelerative forces which occur in experiments involving the stimulation of the semicircular canals has been one of the major stumbling blocks for all investigators in this field (22). As a matter of fact no studies known to the authors have given more than an extremely rough estimate (usually an average) of the accelerative forces involved; in fact, some authors disregard this factor completely, describing the rates of rotation and number of turnings used. These data are relatively useless, since nearly any rate of rotation may be achieved by a wide range of accelerations, and the number of turnings is completely irrelevant, except insofar as there must be enough to permit the effects of acceleration to run their course before decelerative stimuli are introduced. This same limitation applies to this study for the faster rate of rotation, but an

accelerometer[2] was available which gave a measure of the smaller accelerative and decelerative forces at three rpm. The values for acceleration and deceleration presented here were derived from photographic records of four trial runs taken following the completion of the habituation series.

The accelerative force at three rpm built up to a maximum and then dropped off to zero again, giving a photographic record which had the general appearance of a sine curve. The mean duration of the accelerative forces was 1.7 sec., and its maximum value was approximately 13°/sec.[2] The decelerative force showed a very complex pattern although its mean duration was 1.8 sec., being almost the same as the acceleration time. A typical curve exhibited the following sequence of forces within the 1.8-sec. interval: a decelerative force increasing sharply to 104°/sec.[2]; a drop to 20°/sec.[2]; an increase to 52°/sec.[2]; a drop to an *accelerative* force of 48°/sec.[2]; an increase to a *decelerative* force of 46°/sec.[2]; and finally a return to zero. It is obvious from this description that the accelerative forces were very different from the decelerative forces. This is particularly significant in view of the results obtained. At 20 rpm the maximum velocity was attained within approximately 5.5 sec. producing an *average* acceleration of 131°/sec.[2]. Deceleration from 20 rpm to zero occurred in less than 0.5 sec. for an *average* angular deceleration of the order of 240°/sec.[2].

The subjects sat in the Link Trainer with the head inclined forward approximately 25° while they observed the box. Head movements were minimized by a biting board. The subjects reported the onset and offset of the apparent motion (oculo-gyral illusion) of the box, both during rotation and following deceleration. The lag to the first report of motion was so brief, however, that the experimenter counted time from the onset of rotation. During the first series of trials the subjects were turned for 120 sec. for each rate of rotation from onset or until the effects has disappeared, if they persisted beyond two minutes. This gave ample time for the effects of acceleration to reach a negligible value (12, 21). At the end of this period the power was cut off, and the brake was applied simultaneously to produce rapid deceleration. The subject then observed the target for 120 sec. or until the effects had reached zero if they persisted beyond two min. The room lights were turned on for from 15–30 sec. between trials. The subjects' reports of the duration of the apparent motion of the target were timed by the experimenter to the nearest second.

On each of the first 10 experimental days 10 accelerations and 10 decelerations were run for each of the two rotation rates. On the initial day of the test, the two rates of rotation were staggered in a modified *abba* sequence. On succeeding days, however, each subject had 10 successive trials at three rpm and then each took 10 successive trials at 20 rpm after a lapse of from one to three hours. On the eleventh experimental day the procedure was modified to duplicate as nearly as possible the procedure used by Dodge (3) insofar as the number and strength of the stimuli were concerned. The procedure used was the same as that in the foregoing section for the rotation at 20 rpm, except that 19 trials were given in succession; the subjects were turned only 5½ times to the right before deceleration; and no rotations at three rpm were used. This procedure was followed on the four following days in an attempt to determine whether a larger number of strong stimuli per day and the attenuation of decelerative effects by the immediately preceding acceleration might produce habituation.

The three subjects[3] used in this experiment had made many similar observations in this device and in flight. However, they had not been rotated in the device for over a year, and therefore they appeared to be well suited for a study of habituation effects, since a previous study of the habituation or nystagmus has indicated that the effects of habituation, when they occur, dissapear well within that time (21).

Results

The pertinent results are presented in Table I. The apparent motion (oculo-gyral illusion) of the target reported by the subjects was of the same general nature as has been found in other studies (12, 13). At the onset of rotation the target box appeared to move in the direction of rotation. This has been designated as the first effect. Immediately following the first effect the target appeared to move opposite to the direction of turn; this is called the second effect. For one of the subjects on

[2] This instrument was designed and constructed by the Hathaway Instrument Company, Denver, Colorado.

[3] The subjects were the authors and Richard C. Snyder PhM 1/C USN who also assisted with certain technical details.

TABLE I

MEAN DURATION OF THE OCULO-GYRAL ILLUSION IN SEC. FOR THE THREE SUBJECTS FOLLOWING THE VARIOUS ACCELERATIVE AND DECELERATIVE STIMULI

Duration of the first effect

Day	Accel. to 3 rpm	Decel. from 3 rpm	Accel. to 20 rpm	Decel. from 20 rpm	Final Modified Series
1	40.4	34.8	30.3	33.3	29.4
2	47.3	34.7	36.5	30.7	29.2
3	37.8	34.4	33.8	30.3	27.2
4	42.7	36.5	37.0	31.3	29.9
5	40.0	37.6	40.6	34.3	
6	39.5	38.3	36.9	32.2	
7	41.2	36.2	36.7	31.4	
8	39.0	35.6	35.4	32.6	
9	38.4	35.3	37.2	31.2	
10	37.2	32.9	37.2	30.3	

Total duration of the effect

1	61.5	56.6	84.8	84.5	47.2
2	70.9	67.3	77.1	71.5	49.9
3	65.5	66.2	70.7	68.8	44.4
4	67.1	63.8	71.4	69.0	49.1
5	60.3	61.0	74.8	71.7	
6	61.8	67.2	71.0	67.8	
7	60.7	61.3	70.2	67.7	
8	59.8	60.0	72.3	62.9	
9	61.6	55.3	70.1	66.5	
10	60.2	56.7	76.9	63.8	

some trials following the second effect, the target appeared to move in the direction of the turn again; this has been termed the third effect.

Certainly the marked habituation reported by other investigators was not obtained in this study, since the first effect was reported by all of the subjects in all of the trials. The first effect was clearly defined, and the end points were relatively easy to establish. The second effect was reported following acceleration and deceleration from 20 rpm for all trials for subjects C and S, while subject M reported it only rarely and described the effect as 'subjectively vague.' During the trials at three rpm subject S reported the second effect during practically every trial; subject C reported it about 35 percent of the time; while subject M reported it only rarely. The third effect was reported only by subject S; he reported it only following deceleration and observed it in practically all of these trials. Since the first effect was very stable and easy to report, while the second and third effects were highly variable, the analysis of the data has been made in terms of the duration of the first effect and for the total duration of the apparent motion for each trial.

The results for the first series of trials during the initial 10 days show no trends toward reduction in the effect on successive days (Table I) nor were there any tendencies toward reduction within a given series of trials. Considering the first effect alone, there was no change for either of the deceleration series (Table I); there was in fact a slight increase in the duration of the effect for the stronger acceleration and a small decrease in duration for the smaller accelerative stimuli. Essentially the same result was observed for the total duration of all of the effects for each trial. No change was observed in the total duration at three rpm, while a slight decrease was evident at 20 rpm. Taking all of the data into account, the changes were neither clearly defined nor regular for either of these measures of the duration of the effect. In other words, habituation did not occur, either in reduction of frequency or duration of the effects.

The data obtained during the four final experimental days, in which the decelerations followed the accelerations closely in time, show no indication of any tendency toward habituation from day to day nor any change within the 19 trials on any single day (Table I). This was equally true of both the first effect and the total duration of the apparent motion of

the target. The first effect showed a small consistent decrease in duration from the final trials of the first series, whereas the total duration of the effect decreased to a marked degree (Table I). This decrease can be attributed to the fact that the effects of acceleration were still present when the deceleration began. This caused an interference of one effect on the other. It is interesting to note that the largest decrease occurred in the case of the second and third effects rather than in the first effect.

In addition to the lack of any marked objective evidence of habituation of the oculo-gyral illusion, the subjects reported no noticeable difference in the rate of apparent motion of the target or in the general nature of the effects. The subjects did report, however, that mild sensations of discomfort from the experimental situation did decrease and disappear during the early part of the series.

Discussion

It has already been noted that early studies produced ample evidence to show that the nystagmus time could be reduced by repeated rotary acceleration under certain conditions which occurred in the experiments, and that the nystagmus test had little or no value for the selection of aviators. Subsequently, the investigators were concerned with the problem of the modifiability of nystagmus as a basic reflex. However, on the basis of the demonstrated modifiability of the nystagmus reflex, the belief was sometimes expressed that repeated rotary acceleration necessarily caused a reduction in nystagmus and that all other effects also became habituated (17, 19). This point of view was expressed in an excellent historical review of vestibular equilibrium in 1922 by Griffith (19) who states: ". . . It has been demonstrated that all of the effects of ampular stimulatation are highly modifiable under repetition, . . . even by limited repetition in some cases."

In the following decade, however, enough negative evidence had been found to lead Mowrer (35) to a more adequate conclusion:

> Although the negative findings . . . do not in any sense prove that vestibular nystagmus may not, under certain conditions, be considerably reduced or even abolished by repeated elicitation, they do, however, clearly show that this response is not *invariably* diminished by repetition.

Mowrer points out further:

> An analysis of the experimental conditions employed by the various investigators in this field has revealed that a variety of factors may be influential in determining the rate and extent of nystagmus reduction produced by any given schedule of repeated rotation. Different investigators have allowed these factors to operate in their respective experiments in a wide variety of controlled (and uncontrolled) ways, with the result that inconsistencies and contradictions in previously published data have frequently obscured genuine cause-and-effect relationships and have made an understanding of this 'habituation' phenomenon and the conditions conducive to its occurrence extremely difficult.

Mowrer's point of view was substantiated later by one of his own studies, in which he failed to find habituation in 10 human subjects after an extensive series of rotary accelerations and decelerations (38).

The significance of Mowrer's emphasis on the wide variety of controlled and uncontrolled factors which may be involved in the determination of habituation can be seen by examining some of these experimental variables. A few of the more significant factors are:

1. The stimulus, angular acceleration, has frequently not even been mentioned as such, and to our knowledge has never been even approximately measured in earlier studies. The limiting factor is to be found in the fact that no devices were available to measure angular acceleration. Estimates of the angular acceleration may be grossly in error as shown by the fact that the authors initially estimated the stimulus in this experiment to

be an average of 36°/sec.² for both acceleration and deceleration at three rpm. It is clear that this is a gross deviation from the true stimulus situation (see above). It is also evident that the description of the stronger stimuli given above suffers from this same limitation and is not very helpful in describing the stimulus. However, apparatus soon to become available at the Naval School of Aviation Medicine and Research will make it possible to measure angular accelerations over a wide range. An additional variable directly related to the stimulus is whether the head is fixed. Dorcus has already indicated that this must be taken into account (4). The reason is quite clear since the angular acceleration which affects the semicircular canals may vary widely from that of the rotating platform if the head is not fixed.

2. The nature of the organism may have a marked influence on the phenomenon. It is insteresting to note in this connection that most of the data on habituation have been obtained from sub-human organisms. In some of these animals ocular nystagmus occurs, but in the case of pigeons, which have been extensively studied by Mowrer (32–37), head nystagmus is present. The head nystagmus adds a complicating factor in that these movements themselves act as a stimulus to the semicircular canals.

3. Vision during and following the accelerative stimulus has a marked influence on the nystagmus time (21, 38). There is also some evidence that vision may have some effect upon the habituation itself since Holsopple showed that when fixation was eliminated with a +20D. lens, following a series of rotations, the nystagmus time increased (21).

4. One period of acceleration or deceleration may 'damp' a succeeding period if the interval between them is brief, because the after effects of stimulation extend well beyond the period of stimulation. Although the data from this study show that the effects of acceleration may have an effect on decelerative phenomena, as has been shown by others, the effects on habituation itself are unknown. In any event this factor complicates direct comparison between studies.

5. Some accompaniments of rotary acceleration appear to be generally subject to habituation, while others do not. It is generally agreed, for example, that nausea and certain vasomotor changes show very marked habituation, even to the point of complete disappearance, whereas equilibrium and the illusion of turning exhibit practically none. It would appear that nystagmus itself stands some place between the extremes, in that the habituation may occur in varying degrees depending upon factors as yet only partially defined.

A consideration of these variables which have operated in the several experiments make it clear that variability of results is not surprising. It makes it quite clear why Holsopple could remark (22), ". . . One of the most characteristic features of all reported nystagmus measurements is the lack of uniformity of results among different investigators."

In the light of the foregoing discussion an attempt may be made to determine the relation of the current study of the oculo-gyral illusion to the general problem of habituation to rotary acceleration. The findings of this investigation pose certain problems, because habituation of the oculo-gyral illusion was *not* demonstrated, although this illusion has been generally considered to be caused directly by ocular nystagmus, (13, 16, 19, 34) which, as previous investigations have shown, *does* habituate. Considering the nystagmus itself first, five carefully controlled studies of the effects of rotary acceleration upon ocular nystagmus in humans may be cited. Dodge (3) and Griffith (16) using a total of 17 subjects reported virtual abolition of post-rotary nystagmus in some subjects. Griffith's 16 subjects showed an average reduction of 79 percent and the author was of the opinion that, with sufficient

time, 100 percent reduction would have been obtained. Bentley and Dunlap (22) and Holsopple (21) reported a reduction of approximately 50 percent for nine subjects. Holsopple found no drop after the tenth day, although the study continued well beyond that period. Mowrer (38), on the other hand, showed virtually no habituation with 10 male university students. It is significant with regard to Mowrer's investigation that there was absolutely no habituation during darkness, a condition which is more nearly comparable to the procedure used in the present investigation of the oculo-gyral illusion. Both of these studies are in opposition to Griffith's study, which showed a reduction in 'visual vertigo' and nystagmus (16). However, since Griffith's subjects reported only the preliminary 'jerky movements' of the visual field and observed the target on a complex ground, the data cannot be considered to be directly opposed to the present investigation.

The basic purpose of this study was to determine the presence or absence of habituation in the case of the oculo-gyral illusion. Therefore, it was decided to avoid complicating the observations by attempts to determine the occurrence of other factors such as feelings of rotation and nystagmus. This places a limitation upon any attempt to relate the oculo-gyral illusion directly to nystagmus. However, two hypotheses appear to be tenable. First, it appears possible in the light of Mowrer's negative data that the nystagmus time did not change in these subjects, and as a result the oculo-gyral illusion remained constant throughout the experimental runs. A second hypothesis would be that the nystagmus decreased in amplitude to some degree, but that factors such as feelings of turning and changes in bodily pressures, which do not habituate, may have contributed to the illusion to a sufficient degree to leave the duration of the oculo-gyral illusion unchanged. Since there is no evidence that these other factors do affect the illusion, the first hypothesis appears to be the most acceptable.

It should be emphasized that this study gives no direct evidence regarding the habituation of nystagmus, nor does it indicate that the oculo-gyral illusion may not be subject to habituation. It does show clearly, however, that subjects may be exposed to an extensive series of rotary accelerations without producing any marked effect on this illusion, and therefore supports the incidental findings of earlier studies of the oculo-gyral illusion which found no habituation (1, 2, 12, 13, 25).

With regard to the contribution of these findings to visual orientation in flight, it can be said that this illusion of orientation may be expected to occur in pilots of aircraft at night when rotary accelerations above threshold occur. The flight experience of the pilots is probably not an important factor. This can be inferred on the basis of the lack of habituation observed in this study, and the fact that by no means are pilots in general necessarily subject to strong and repeated rotary accelerations in flight. The number of such strong stimuli affecting pilots flying multi-engine transport aircraft, for example, would probably be much less than those occurring in this experiment. In smaller aircraft the possibilities of repeated rotary accelerations would be determined by the nature of the operations involved in the flight. When exact measurements of the actual angular accelerations occurring in specific flight ma-

neuvers are obtained, important data on this point will be available.

Summary

The effects of repeated angular acceleration and deceleration on visual perception were studied in three subjects to establish whether any change in the oculo-gyral illusion would occur. The subjects were rotated while they sat in a Link Trainer which rotated only about a vertical axis. The subjects were turned for 10 successive trials at three rpm and then for 10 successive trials at 20 rpm on 10 successive days. On the following four days the subjects took 19 successive trials at 20 rpm. Measures of the angular accelerations and decelerations were available at the slower rotation speeds. The subjects observed a complex target in the dark and reported the onset and offset of the oculo-gyral illusion following both acceleration and deceleration for each trial. The duration of the effects was recorded in an adjacent room by the experimenter. The three subjects were familiar with the effects observed, but they had not been rotated for approximately one year, so they were suitable for a study of habituation.

No evidence of habituation was found for any of the effects observed. Immediately following acceleration and deceleration the subjects reported apparent motion of the target (oculo-gyral illusion, first effect). They also reported a reversal of this motion when the first effect ended; on some trials one subject reported a third effect of brief duration following the second effect. A detailed analysis of the three effects showed no habituation for any one of them nor for the total duration of the apparent motion. The results also support the hypothesis that the nystagmus remained unchanged. Although this study does not preclude the possibility of habituation of the oculo-gyral illusion under other conditions, it is clear that habituation does not necessarily follow a series of strong rotary accelerations and decelerations.

(Manuscript received May 26, 1948)

References

1. Clark, B., Graybiel, A., & MacCorquodale, K. The illusory perception of movement caused by angular acceleration and centrifugal force during flight. II. Visually perceived motion of a fixed target during turns. *J. exp. Psychol.*, 1948, **38**, 298–309.
2. Clark, B., Graybiel, A., & MacCorquodale, K. Illusory rotation of a target during turns in an aircraft. *Amer. J. Psychol.*, 1948, **61**, 50–58.
3. Dodge, R. Habituation to rotation. *J. exp. Psychol.*, 1923, **6**, 1–35.
4. Dorcus, R. M. The effect of intermittent rotation on orientation and the habituation of nystagmus in the rat and some observations on the effects of pre-natal rotation on post-natal development. *J. comp. Psychol.*, 1933, **15**, 469–475.
5. Dunlap, K. The nystagmus test and practice. *J. Amer. med. Assn.*, 1919, **73**, 54–55.
6. Dunlap, K. Adaptation of nystagmus to repeated caloric stimulation in rabbits. *J. comp. Psychol.*, 1925, **5**, 485–493.
7. Fearing, F. The retention of the effects of repeated elicitation of the post-rotational nystagmus in pigeons. I. The retention of the effects of massed stimulation. *J. comp. Psychol.*, 1940, **30**, 31–40.
8. Fearing, F. The retention of the effects of repeated elicitation of post-rotational nystagmus in pigeons; retention of effects of distributed stimulation. *J. comp. Psychol.*, 1941, **31**, 47–56.
9. Fearing, F., & Mowrer, O. H. The effect of general anesthesia upon the experimental reduction of vestibular nystagmus. *J. gen. Psychol.*, 1934, **11**, 133–144.
10. Fisher, L., & Babcock, H. The reliability of the nystagmus test. *J. Amer. med. Assn.*, 1919, **72**, 779–782.
11. Grammer, G. The static equilibrium of airplane pilots. *J. comp. Psychol.*, 1925, **5**, 345–364.

12. Graybiel, A., & Hupp, D. I. The oculogyral illusion: a form of apparent motion which may be observed following stimulation of the semicircular canals. *J. aviat. Med.*, 1946, **17**, 3–27.
13. Graybiel, A., Clark, B., MacCorquodale, K., & Hupp, D. I. The role of vestibular nystagmus in the visual perception of a moving target in the dark. *Amer. J. Psychol.*, 1946, **59**, 259–266.
14. Graybiel, A., Hupp, D. I., & Patterson, J. L., Jr. The law of the otolith organ. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **35**, 35.
15. Griffith, C. R. The decrease of after-nystagmus during repeated rotation. *Laryngoscope*, 1920, **30**, 129–137.
16. Griffith, C. R. The organic effects of repeated bodily rotation. *J. exp. Psychol.*, 1920, **3**, 15–46.
17. Griffith, C. R. An experimental study of dizziness. *J. exp. Psychol.*, 1920, **3**, 89–125.
18. Griffith, C. R. Concerning the effect upon nystagmus of repeated rotation. *Laryngoscope*, 1920, **30**, 22–25.
19. Griffith, C. R. An historical summary of vestibular equilibrium. *Univ. Ill. Bull.*, 1922, **20**, No. 5, 1–178.
20. Henderson, J. W. Optokinetic and other factors modifying vestibular nystagmus. *Arch. Ophthal.*, 1947, **37**, 459–471.
21. Holsopple, J. Q. Some effects of duration and direction of rotation on post-rotation nystagmus. *J. comp. Psychol.*, 1923, **3**, 85–100.
22. Holsopple, J. Q. Factors affecting the duration of post-rotation nystagmus. *J. comp. Psychol.*, 1923, **3**, 283–304.
23. Holsopple, J. Q. An explanation for the unequal reductions in post-rotation nystagmus following rotation practice in one direction. *J. comp. Psychol.*, 1924, **4**, 185–193.
24. Holsopple, J. Q. Space and the non-auditory labyrinth. In: Murchison, C. *Foundations of experimental psychology*. Clark Univ. Press, Worcester, 1929, pp. 414–433.
25. Kerr, W. A., & Graybiel, A. Thresholds of stimulation of the horizontal semicircular canals in man. *Amer. Psychologist*, 1946, **1**, 237–238.
26. Lewis, E. R., & Pike, F. H. The Barany chair test and flying ability. *J. Amer. med. Assn.*, 1918, **70**, 1559–1560.
27. Loch, W. E., & Haines, H. L. Habituation to caloric stimulation. *Laryngoscope*, 1946, **56**, 43–45.
28. MacCorquodale, K. The effects of angular acceleration and centrifugal force on non-visual space orientation during flight. *J. aviat. Med.*, 1948, **19**, 146–157.
29. MacCorquodale, K., Graybiel, A., & Clark, B. The effects of angular acceleration and centrifugal force on non-visual space orientation during flight. II. Influence of habituation and technique of assuming the turn. BuMed. Res. Project X-148(Av-4-3) Report No. 15, 22 July 1946.
30. Maxwell, S. S., Burke, V. L., & Reston, C. The effect of repeated rotation on the duration of after-nystagmus in the rabbit. *Amer. J. Physiol.*, 1921, **58**, 432–438.
31. Maxwell, S. S. The effect of habituation on the rotation-nystagmus as compared with the after-nystagmus in the rabbit. *Amer. J. Physiol.*, 1924, **61**, 125–126.
32. Mowrer, O. H. Concerning the normal functioning of the vestibular apparatus. *Ann. Otol. etc.*, 1932, **41**, 1–10.
33. Mowrer, O. H. The influence of "excitement" on the duration of post-rotational nystagmus. *Arch. Otolaryn.*, 1934, **19**, 46–54.
34. Mowrer, O. H. An analysis of the effects of repeated bodily rotation with especial reference to the possible impairment of static equilibrium. *Ann. Otol., etc.*, 1934, **43**, 367–386.
35. Mowrer, O. H. The modification of vestibular nystagmus by means of repeated elicitation. *Comp. Psychol. Monogr.*, 1934, **9**, No. 5, 1–48.
36. Mowrer, O. H. Some neglected factors which influence the duration of post-rotational nystagmus. *Acta otolaryng.*, 1935, **22**, 1–23.
37. Mowrer, O. H. "Maturation" vs. "learning" in the development of vestibular and optokinetic nystagmus. *J. genet. Psychol.*, 1936, **48**, 383–404.
38. Mowrer, O. H. Influence of vision during bodily rotation upon the duration of post-rotational vestibular nystagmus. *Acta otolaryng.*, 1937, **25**, 251–263.
39. Parsons, R. P., & Segar, L. H. Barany chair tests and flying ability: A correlation study of 100 naval aviators. *J. Amer. med. Assn.*, 1918, **70**, 1064–1065.
40. Pilz, G. F. On the relation of after-nystagmus to rotation nystagmus. *Amer. J. Physiol.*, 1926, **77**, 428–458.
41. Yacorzynski, G. K., Halstead, W., & Fearing, F. Relationship between the experimental reduction of vestibular nystagmus and equilibrium. *J. Psychol.*, 1941, **11**, 161–171.

THE EMERGENCE OF A TONAL SENSATION [1, 2]

BY J. DONALD HARRIS AND CECIL K. MYERS

Medical Research Laboratory, U. S. N. Submarine Base, New London, Conn.

I. Introduction

When an absolute threshold of auditory acuity is obtained for a pure tone, the intensity is usually decreased stepwise from some easily heard level to well below threshold, whereupon the intensity is increased in steps until the tone is again heard clearly. The average dial reading, to the nearest 5-db setting, of several crossings of the threshold, is usually entered as the absolute threshold for that tone (1). The rationale of using 5-decibel steps, to which all commercial audiometers now conform, was simply clinical convenience, plus the observation (13) that the accuracy of the average audiometrician is only slightly less than ± 5 db.

It is, however, quite possible to obtain precision at least twice as good as this even with the clinical audiometer (6), so that it would seem desirable in the laboratory to test auditory acuity using steps of two or even of one db.

Now when the acuity of an ear is carefully examined by the use of these finer steps, with many stimuli presented at or very close to threshold, a phenomenon occurs which is ordinarily overlooked when larger intensity steps are used: a change in the *quality* of the tone seems to occur as the stimulus intensity increases. For example, at an intensity such that the stimulus is reported only 30 percent of the time, the tone when heard has an indefinite pitch; whereas at an intensity several db above threshold, the tone takes on a 'sharp' quality and is clearly perceived as a pure tone.

The suggestion is made by Dr. Hallowell Davis (3) that the zone between the pure tone intensity at which 'something' can be detected 50 percent of the time, and the intensity at which a 'pure tone' is reported 50 percent of the time, be given the name 'zone of detectability.'

In a previous report (9) the present authors studied the zone of detectability as a function of frequency from 500 through 14,000~, compared the slopes of the psychophysical function for the detection threshold and for the pure tone threshold, and noted the effects, on the zone of detectability and on the slopes of the functions, of introducing a white noise of various intensities. They concluded:

> We do not suppose that there is any *sharp* emergence of pitch-quality as a tone is increased from subthreshold strength to audibility and above, nor do we suppose that, once a pitch-quality can be ascertained 50% of the time, there is no further refinement of the pitch-quality as the tone becomes even louder. This is probably a fairly gradual process. Our own present purpose was served when we had determined the general intensity-range above detectability where a pure tone could be recognized as having a pitch. No doubt a careful observer could use criteria which would enable him to fix one or more thresholds *in addition* to the two studied here.

The existence of an area or a zone within which pure tones might be heard but not identified as such, has obvious significance in many life situ-

[1] Opinions or conclusions are the authors', and do not necessarily reflect the official views of the Navy Department.

[2] A preliminary report was presented before the Easton Psychology Association, 25 April 1947. See: Myers, C. K., & Harris, J. D. Detection thresholds and tonal thresholds in auditory acuity. *Amer. Psychologist*, 1947, **2**, 433.

ations. Furthermore, relations appear between the phenomenon and theories of threshold fluctuation, of integration of sound energy, of the stages culminating in a true tonal sensation, and of the relation of tonal threshold to pitch discrimination.

The present article extends the original data, without any duplication, and in addition compares material on the zone of detectability with data on pitch-matching at the same sensation levels.

II. Apparatus

For our purposes the commercial audiometer was too coarse and unstable an instrument. Our assembled equipment is shown in Fig. 1. We used two General Radio 913c beat-frequency oscillators, and calibrated for frequency against a primary frequency standard. The outputs were led through attenuators with steps of 0.1 db to the two inputs of a combined electronic switch and fader amplifier.[3] This system passed a first tone, serving solely as a warning signal, for 330 msec.; allowed one sec. of silence; and then passed a second tone, the experimental stimulus, for 330 msec. This cycle was repeated every five sec. Any tone required 100 msec. to build up to maximum, and 100 msec. to decay.

The output of the amplifier was then led to an attenuator with one-db steps, an impedance-matching transformer, and a single headphone, Permoflux Type PDR-8. The phone was mounted in a strong spring headband and an oval cushion of spongy material, Service Type NAF-48490-1. The non-used ear was covered with a dummy phone and cushion. These cushions in tight headbands provide attenuation of 10–30 db for pure tones of the frequencies we used.

Purity of the experimental stimulus was satisfactory as determined with a cathode ray oscilloscope and with a General Radio wave analyzer Type 736-A. Switching transients were suppressed more than 40 db below audibility when the stimulus was at threshold intensity. The experiment took place in a soundproof suite with attenuation from the outside sufficient to bring all noises in or around the building to inaudibility. Apparatus and *S* were in separate rooms, with two quadruple gasket doors between. Communication between *S* and *E* was by way of a two-way phone which carried sound to the *S* only when the *E* pressed a button.

The headphone was first calibrated with a Ballantine Model 505 6-cc artificial ear. From

[3] Design furnished through the courtesy of Dr. John E. Karlin, Bell Telephone Laboratories, and kindly constructed for us by Mr. Harold Berggren in the Electronics Shop of U. S. Naval Underwater Sound Laboratory.

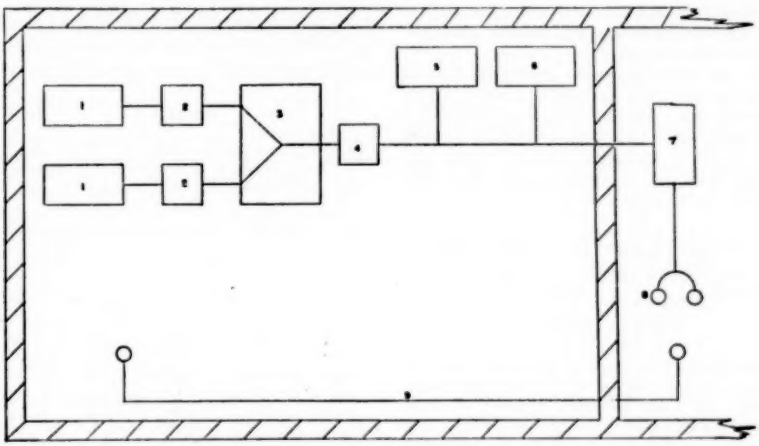


Fig. 1. 1. General radio oscillator, type 913 C; 2. Attenuator in .1 db steps; 3. Electronic switch and amplifier; 4. Attenuator in 1.0 steps; 5. Dumont cathode ray oscilloscope; 6. General radio wave analyzer, type 736 A; 7. Ballantine vacuum tube voltmeter, model 300; 8. Permaflux headphones, PDR 8; 9. Two-way speech communication.

these data one can calculate the sound pressure developed in the ear for any voltage across the headphone. The voltage across the headphone at each threshold intensity was determined with a Ballantine decade amplifier and vacuum tube voltmeter, and converted into sound pressure *re* 0.0002 dyne/cm.^{2}.

III. Method of Determining Thresholds

To determine 'detection' thresholds and 'tonal' thresholds a modification of the method of limits was used. The frequency for a certain experimental session was selected and presented as a stimulus with the second oscillator (the first oscillator merely served as a signal to the *S* that a weak sound would follow one sec. later). The approximate threshold of the test frequency was found, and test items were presented in one-db steps at random around this value. The *S* was required to judge and report by phone whether he heard 'nothing,' 'something,' or a 'tone' during the relevant time interval.

By this method two thresholds are simultaneously found, namely, the 50 percent transition point, from 'nothing' to 'something,' and the 50 percent transition point from 'something' to a 'tone.' An alternate method, presenting the intensities in gradually ascending or descending order instead of at random, was discarded almost immediately, because early in the session the width of the transition zone very soon was known to the *S* and was regularly influencing decisions later on in that session. The random method of presentation avoided this complication to a large extent.

An experimental session continued until 10 judgments had been recorded at every intensity level from that where the *S* reported 100 percent 'nothing' to that where he reported 100 percent 'tone.' The best-fitting ogives were then plotted, and the 50 percent points and their probable errors calculated.

IV. Experiment I: Establishing the Difference between Thresholds ('Zone of Detectability')

Procedure.—Since trained *S*s were essential for this experiment (the writers served alternately as *S* and *E*) it was necessary to avoid so far as possible the effect of knowledge of results. Accordingly, in a first experiment *E* placed the test frequency well below threshold, and increased it gradually but irregularly to the intensity at which *S* began reporting the presence of some tonal quality; but the tone was never raised much above this intensity, so that in an experimental session *S* never experienced the clear cut tonal quality of an intense pure tone. Thresholds were thus collected on successive days until 10 or more were obtained at each of the frequencies 500, 1000, 2000, 4000, 8000, and 14,000~.

Results.—The existence of a true 'zone of detectability' depends upon there being a statistically reliable separation, in decibels, of the threshold of tonality from the threshold of detectability. This will depend in turn upon the probable errors of the thresholds themselves.

It must be said at the start that the threshold of tonality is every bit as reliable as the threshold of detectability, either by the criterion of the probable error of the individual threshold, or by the criterion of the standard deviation of successive thresholds. It might perhaps be thought that the distinction between 'nothing' and 'something' would be sharper than the distinction between 'something' and 'pure tone'; but the case is not rare when indeed the threshold of tonality exhibits an even steeper psychophysical function than its companion threshold of detectability.

Table I presents the mean differences between the two thresholds, and

TABLE I

Mean Differences Between Thresholds of Detectability and Companion Thresholds of Tonality in Decibels

	CKM		JDH	
	Mn. Diff.	1 S.E.	Mn. Diff.	1 S.E.
500	2.0	.18	3.0	.29
1,000	1.7	.19	3.1	.27
2,000	2.3	.23	2.7	.23
4,000	1.8	.14	3.5	.24
8,000	3.2	.31	5.1	.59
14,000	2.9	.27	4.9	.50

Range of probable errors of individual thresholds: CKM .23–.59 db; JDH .10–.77 db.

Each mean difference is an average of 10 or more differences.

their standard errors. Each mean is the average in decibels of 10 or more differences. The difference between any particular companion pair of thresholds is statistically reliable.

V. Experiment II: Selecting and Maintaining a Criterion of 'Tonal Quality'

Procedure.—The previous experiment confirms a subjective impression that when a tone is raised from inaudibility to audibility, a rather loosely-fixed intensity exists at which *S* is willing to report the presence of a definite tonal quality. To throw more light on this point, a somewhat different procedure was next undergone. In a second experiment, *S* was required to choose a more rigid criterion of tonality: instead of reporting 'tone' whenever the stimulus had *any* (even though nebulous) pitch quality, he was to report 'tone' only when the stimulus seemed to have a rather definite pitch such that it could be at least roughly discriminated from other pure tones. The object of this experiment was to determine whether the *S* could select and maintain a criterion with any degree of stability.

The procedure of this experiment was exactly the same as that of the first except that the *S* knew what frequency was being presented, and he was free to ask for an increase in intensity of tone at any time before or during the presentation of the test items. It was thought that by this procedure the *S* need not necessarily ever be at a loss to know the general level of his criterion.

Results.—The results of this procedure are indistinguishable from those of the first procedure except at the higher frequencies. Here the zone of detectability becomes wider for both *S*s. For *S* No. 1, at 8000 and 14,000~, the zone increases to nearly 4 db, while for *S* No. 2 it increases to 6 and 8 db respectively.

These data indicate that it is quite possible for *S* to select a certain intensity of tone at which a certain quality of tonality seems to him to be present, and to maintain that criterion when the intensities are shifted in a random manner.

VI. Experiment III: Pitch-Matching as Evidence of Tonality

Procedure.—In order to provide an objective measure of the pitch-quality of weak tones, *S* was asked to match that tone in pitch with another fixed tone of distinct pitch-quality. The pitch-quality of the weak tone was measured by the accuracy of frequency matching.

The same apparatus was used as before. *S* sat in front of the oscillators and the switch, which has a silent manual left-right key turning either oscillator on for any length of time, or off at will. The threshold of detectability for the frequency to be used was carefully determined by the method of limits at the beginning of each experimental session. Then, the fixed-frequency oscillator output was set 10 db above that value and the variable-frequency oscillator set either 0, 1, 2, 3, 4, 5, 6, 8 or 10 db above that value. *S* then closed his eyes, whereupon *E* turned the frequency dial of the variable oscillator at random, and *S* was asked to turn that dial[4] until by switching left and right at will he judged the two tones to be matched in pitch. The psychophysical procedure described here is the method of average error, and the measure of precision of matching is taken to be the standard deviation of all matching errors regardless of sign. The magnitude of experimental error is taken to be the standard error of this standard deviation. These values were accordingly calculated for each frequency at every intensity level of the test frequency. In addition, record was kept of whether each error of matching was positive or negative, so that any systematic over- or under-estimation of pitch could be calculated.

Results.—The results of pitch-matching for each of the six frequencies are contained in Table II. Standard deviations and their standard errors are given.

In order to provide a quick picture symbolizing the way in which a tone increases in sharpness of pitch-quality, Fig. 2 shows how the definiteness of pitch-matching increases as the tone increases in intensity. The standard deviations are laid out each around an arbitrary origin and are enclosed by freehand curves. The freehand

[4] In addition to the main tuning dial, *S* had available a second dial calibrated in steps of one ~, 50~ in 150°.

TABLE II

STANDARD DEVIATION OF ERRORS IN PITCH-MATCHING (CKM) IN CYCLES PER SEC.
Level Above Threshold of Detectability of Variable Frequency
(Fixed frequency set 10 db above that threshold)

		10	8	6	5	4	3	2	1	0
500	S.D.	1.9	3.0	3.5	6.0	3.1	5.3	9.0	7.6	18.5
	S.E.	.6	1.0	1.2	1.2	1.0	1.8	2.1	1.8	4.3
1,000	S.D.	3.5	3.3	5.1	5.6	5.7	11.1	12.6	14.0	19.5
	S.E.	1.7	1.1	1.7	1.9	1.9	3.7	4.2	4.7	6.5
2,000	S.D.	2.2	3.2	8.7	4.7	9.4	5.8	18.6	16.2	24.1
	S.E.	.7	1.1	2.0	1.6	3.1	1.4	6.2	3.8	5.7
4,000	S.D.	3.5	7.7	9.0	14.9	15.3	15.8	26.4	30.5	44.0
	S.E.	1.2	2.6	3.0	3.5	4.8	5.3	6.2	10.2	14.7
8,000	S.D.	8.5	41.2	38.8	52.8	52.5	62.5	131.2	158.8	
	S.E.	2.8	13.7	12.9	17.6	17.5	20.8	43.7	52.9	
14,000	S.D.	34.5	41.3	65.7	69.3	98.9	113.0	127.1	92.2	171.0
	S.E.	11.5	8.3	21.9	23.1	33.0	37.7	42.4	21.7	57.0

curves do not approach an asymptote at 10 db, since it seems clear that a further increase in loudness would produce even better discrimination. It must be noted that the base lines for the six frequencies are laid out in three different scales so as to place all the data in a single, simple, figure. If the standard deviations were each divided by the pertinent frequency,

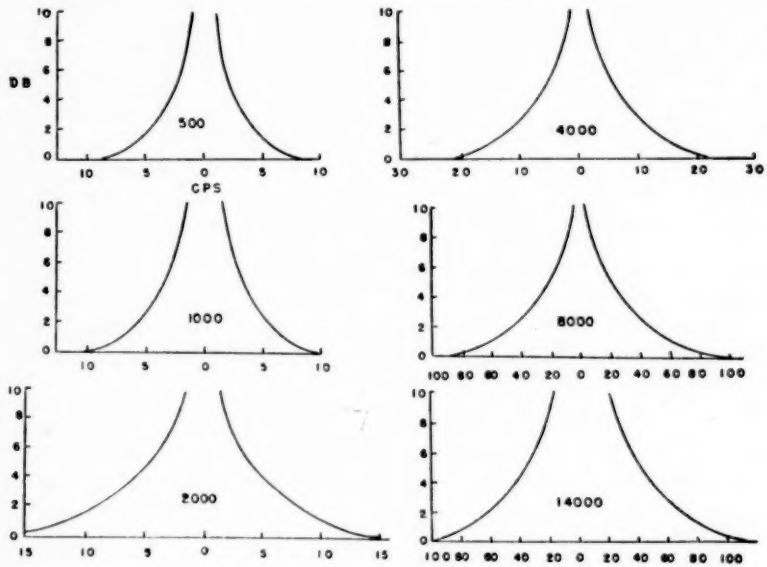


FIG. 2. Symbolizing the emergence of a tonal sensation. (See text.)

the percentage precision would be seen to vary only slightly with frequency.

VII. Discussion

1. *The continuity of tonal sensation.*—As a tone is increased in intensity from inaudibility to well above threshold, the change in tonal sensation is not divided into clear-cut stages. On the contrary, when the analysis is fine enough the change is seen to be a smoothly continuous process. Thus, it must be emphasized that the distinction spoken of between 'something present' and 'pure tone present' has only a verbal sharpness. The distinction is not without meaning, since the probable error of the 50 percent dividing point is satisfactorily small, but in speaking of different thresholds one must not be misled into thinking of them as definite breaks in the continuity of tonal sensation as it emerges from nothing through an amorphous interval to a quite definite pure tone.

This experiment has shown that at a definable intensity above absolute threshold a pure tone begins to take on a tonal quality, and that *S* can select and maintain a criterion of definite tonal quality at a still higher intensity. But we do not suppose that these three thresholds exhaust the possibilities. Probably at least half-a-dozen points instead of these three could be chosen by a careful *S* and maintained with some consistency. On this line of reasoning, it will be remembered that Shower and Biddulph's *S*s (11) were not discriminating frequencies at their optimum until the sensation levels were about 30 db. Presumably, at any intensity below 30 db, some relative indefiniteness of tonal quality exists.

Experiment III constitutes what we regard as crucial evidence on the continuity of emergence of tonal sensation (a concept which would be expected from a consideration of practically every other auditory growth function). While our data extend only to a sensation level of 10 db, it is certain that from absolute threshold to that intensity, the emergence of tonality is perfectly continuous—not, indeed, in straight-line fashion necessarily (we have no way of translating the curves of Fig. 2 into sensation units)—but at any rate not in a series of discrete steps.

2. *Relation to other experiments.*—The average difference, in decibels, between the thresholds of detectability and of tonality is given by Pollack (10) at seven octaves from 125–8000~ inclusive. The data are of the same order of magnitude as we found in our Experiment II by a different psychophysical method, and differences between the two sets of results can easily be explained on methodological grounds. The similarities, indeed, are more striking than the differences. At the same time, we believe it can be shown that the psychophysical method which we adopted would have a slight advantage over that of Pollack in future investigations on the zone of detectability.

It is a question whether such data can justifiably be averaged from *S* to *S*. We have no assurance that all *S*s were using the same criterion. Accordingly, individual zones of detectability from both experiments are plotted in Fig. 3. The reference (zero) level is in each case *S*'s threshold of detection, above which his threshold of tonality may be found for any frequency. (All five *S*s had essentially normal hearing.)

Both experiments agree in there being only a slight effect of frequency. The present data show

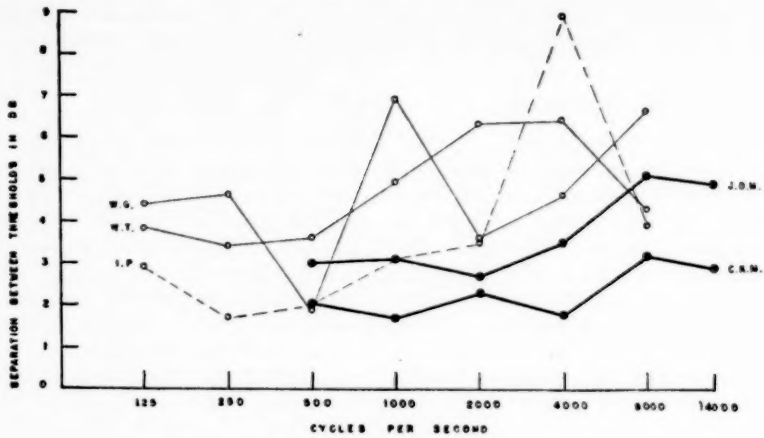


FIG. 3. Zone of detectability

reliable differences for both *S*s between any tone up to and including 4000~ and either of the two higher tones. Among the 63 possible inter-frequency comparisons of Pollack's data (21 for each *S*), W.T. exhibits no statistically reliable difference between any two frequencies; I.P. exhibits one, between 250~ and 1000~ (excluding an anomalously large and statistically unreliable difference at 4000~); and W.G. exhibits five, between 500~ and all other frequencies except 2000~. It would seem premature at the present time to say anything more precise than that for some *S*s the zone of detectability is a little greater for high than for low tones.

At 500~, the zones of detectability for the two experiments are indistinguishable. But at 1000~ and higher, the zones of Pollack's *S*s are in almost every case larger than ours. This difference may almost certainly be explained by reference to the procedures used. Pollack's *S*s were in control of the stimulus intensity by means of a motor-driven attenuator: turning a switch to one side increased and to the other side decreased the intensity. By this method many of the constant errors in the method of average error can be overcome, but note that *S* was at any time able to increase the intensity as loud as he wished. Under these conditions *S* is free to pick and continually corroborate a particular threshold. These conditions are much like our Experiment II. In this connection, recall that the zone of detectability in our Experiment II increased several db over Experiment I for higher tones—just the region of discrepancy between Pollack's data and those of our first procedure. We conclude that the zone of detectability depends to a good extent on the precise psychophysical method used, and particularly on what control *S* can exert, directly or indirectly, on the stimulus. It is our opinion that our first procedure is ideally suited to determine the separation between a threshold of detectability and a threshold of incipient tonality; and that modifications involving stimulus control by *S* may produce data on any clearer level of

tonality which *S* chooses to select and maintain.

3. *Is the criterion one of pitch or of loudness?*—The difficult question must be attacked, whether in selecting a criterion of 'tonal quality' the *S* chooses a certain loudness, or perhaps volume, rather than a certain definiteness of pitch. The first two parts of this experiment do not provide any clue—what they do is document the existence and precision of a certain criterion; that criterion is not necessarily one of pitch.

Two lines of evidence, however, point to the probability of pitch as the major basis of the criterion. First, in Experiment III, *S* matched tones on the basis of pitch, not loudness; and there are certain correspondences between this work and that of the first two parts. Second, Harris (7) showed that pitch discrimination becomes progressively poorer as the intensity of a tone is decreased above a white noise mask, and that discrimination finally reaches a minimum at a point asymptotic not with the 50 percent detectable point, but 2–3 db *above* this point. But this is just the region at which tonal quality is becoming apparent, and the correspondence indicates that the tonal threshold introduced here is in all probability a pitch phenomenon.

A somewhat similar line of reasoning was followed by Doughty and Garner (4), who determined for a loud tone the duration in milliseconds which produced a certain pitch-quality. They found that this duration corresponded roughly to the duration of that same tone which was sufficient for 'effective' pitch discrimination (Turnbull, 12).

4. *Distinction between tonality changes caused by reducing physical duration and caused by reducing intensity.*—Previous studies (2, 4, 5) have shown that the pitch-quality of a tone can be reduced simply by shortening its physical duration. The first question to ask, then, is whether the reduction in pitch-quality found here can likewise be attributed to a shortening of duration, in this case *apparent* duration. That the weaker tones used here were significantly shorter in apparent duration can be inferred from the work of Lifschitz (8) on similar low sensation levels.

But both the physical and psychological effects are quite different when a tone is shortened by physical means and again by reducing its intensity to near threshold. In the first case, what happens is that more and more of the total energy is concentrated in frequency bands *other than* the one desired. At first shortening, a pure tone seems to start with a click and to stop with a click. With further shortening, a very brief pure tone seems to be surrounded by clicks, and finally, the two clicks merge and no pure tone can be sensed. What happens is quite different when a tone is reduced in *apparent* duration by reducing its intensity. Here other frequencies than the one desired are of even *less* significance as duration decreases, since they are of less energy than the experimental tone. No click sensation is experienced, as is predominantly the case with physical shortening.

We conclude, therefore, that the gradual loss of tonality when intensity is reduced is a different sort of phenomenon from the loss of tonality when physical duration is reduced. The latter is a function of stimulus frequency-composition, while the change in tonality reported here is definitely a function of stimulus intensity near threshold.

5. *The relation of loudness and of apparent duration to tonality.*—But

although we cannot use the previous experiments on physical duration and tonality to explain the results of this experiment, nevertheless we cannot overlook a possible contribution of *apparent* duration to tonality. It is possible that a weak tone, rendered apparently very short, may *for that reason* be reduced in tonal quality. It is also possible that tonality is a consequence of the loudness of a tone.

Whether tonality depends upon apparent duration or upon loudness in any direct sense cannot be stated at present. The very low loudnesses dealt with here have never been carefully studied. It is probable that reduced loudness, apparent duration, and tonality are all three merely psychological expressions of the neurological state of affairs at near-threshold intensities, namely, a reduction in number of nerve fibers excited and in number of impulses reaching the central nervous system. One cannot do more at present than indicate the major contribution of intensity rather than of frequency to the emergence of tonality.

Summary and Conclusions

When the absolute threshold of auditory acuity for pure tones is carefully explored with fine intensity steps, a region called the 'zone of detectability' can be noticed by experienced *S*s. In this zone, pure tones can be detected but cannot be said to have a true pure-tone quality. This zone of detectability is defined as the intensity area between a 50 percent detection threshold and a 50 percent pure-tone threshold.

In Experiment I, these two thresholds were simultaneously determined by a modification of the method of limits at each of six frequencies, 500, 1000, 2000, 4000, 8000, and 14,000~. Two experienced *S*s were used. The zone of detectability varied from 2.7 to 5.1 db for one *S* and from 1.7 to 3.2 for another *S*. A slight tendency was found for the zone to be larger at the highest frequencies. In Experiment II, it was shown that particularly for the higher frequencies, an experienced *S* can select a certain criterion of tonality and maintain it with considerable precision.

In Experiment III, *S* matched a variable frequency of very low sensation level to a standard frequency of 10 db sensation level. The frequency-match was progressively better as the intensity of the variable frequency increased from zero to 10 db sensation level, and the course of this improvement was taken to symbolize the gradual nature of the emergence of a tonal sensation from an indefinite 'something' to the sharp sensation of a clear pure tone.

It was reasoned that the subjective criterion of 'tonality' which an *S* can select and maintain is indeed one of pitch rather than of loudness; but it was concluded that the phenomenon depends upon the intensity of the stimulus rather than upon a change in its frequency-composition.

(Manuscript received June 21, 1948)

References

1. Bunch, C. C. *Clinical audiometry.* St. Louis: C. V. Mosby, 1943.
2. Burck, W., Kotowski, P., & Lichte, H. Der Aufbau des Tonhöhenbewusstseins. *Elek. Nachr.-Techn.*, 1934, **12**, 326–333.
3. Davis, H. Personal communication.
4. Doughty, J. M., & Garner, W. R. Pitch characteristics of short tones. I. Two kinds of pitch threshold. *J. exp. Psychol.*, 1947, **37**, 351–365.
5. Ekdahl, A. G., & Stevens, S. S. The relation of pitch to the duration of a tone. Unpublished. Cited in: Stevens, S. S., & Davis, H. *Hearing.* N. Y.: Wiley, 1938, p. 101.

6. HARRIS, J. D. Free voice and pure tone audiometer for routine testing of auditory acuity. *Arch. Otolaryngol.*, 1946, **44**, 452–467.
7. HARRIS, J. D. Studies on pitch discrimination in masking. II. The effect of signal/noise differential. *J. acous. Soc. Amer.*, 1947, **19**, 816–819.
8. LIPSCHITZ, S. Fluctuation of the hearing threshold. *J. acous. Soc. Amer.*, 1939, **11**, 118–121.
9. MYERS, C. K., & HARRIS, J. D. Detection thresholds and pure tone thresholds in auditory acuity. U.S.N. Bureau of Medicine & Surgery Report, Project NM-003-022.
10. POLLACK, I. The atonal interval. *J. acous. Soc. Amer.*, 1948, **20**, 146–149.
11. SHOWER, E. G., & BIDDULPH, R. Differential pitch sensitivity of the ear. *J. acous. Soc. Amer.*, 1931, **3**, 275–287.
12. TURNBULL, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.
13. WITTING, E. G., & HUGHSON, W. Inherent accuracy of a series of repeated clinical audiograms. *Laryngoscope*, 1940, **50**, 259–269.

THE PROBLEM OF STIMULATION DEAFNESS. III. THE FUNCTIONAL AND HISTOLOGICAL EFFECTS OF A HIGH-FREQUENCY STIMULUS *

BY KENDON R. SMITH ** AND ERNEST GLEN WEVER

Princeton University

This paper is one of a series which describes quantitative measurements of the effects of over-stimulation of the ear with sounds (1, 2). Earlier studies used exposure tones of 300, 1000, and 5000 cycles per sec., all delivered at a standard intensity and duration; and here the investigation is extended by the use of 10,000 cycles.

Guinea pigs were used as subjects, and both electrophysiological and histological evaluations made of the acoustic injury.

Procedure

A determination of cochlear function in terms of electrical potentials was first made for the normal condition, before exposure. The measurements were made with 11 tones over the range 100 to 10,000 cycles and consisted of a determination of the sound intensity required at each of these 11 frequencies to produce a standard response of 10 microvolts. Then the ear was stimulated for four min. with a 10,000 cycle tone at a sound pressure of 1000 dynes per sq. cm. Immediately thereafter followed a redetermination of the cochlear potentials. Finally, the ear was prepared for histological study, first by intravital perfusion with a mixture of modified Maximow's fluid and 10 percent formalin, then by decalcification, celloidin embedding, and sectioning at 20 micra. Every fifth section was stained in hematoxylin and eosin and mounted serially. Observations of microscopic structure were plotted according to Guild's (3) reconstruction method, which yields a map of the cochlea to scale; this map became the basis of linear graphs showing the condition of the various structures throughout the cochlea. The study was carried out on 10 ears.

* This investigation was initiated under a grant from the John and Mary R. Markle Foundation, and continued as a part of a contract (N6-onr-27, III) with the Office of Naval Research.

** Now at The Pennsylvania State College.

Results

Electrical Tests

As in the earlier studies, the electrical tests revealed large individual differences in the susceptibility of the ears. Fig. 1 indicates both the nature and degree of variation of the changes. The curves show the sensitivity changes occasioned by the exposure: they represent, in decibels below the normal or zero line, the impairment now present; or, more precisely, the extent to which the stimulus had to be increased to elicit the standard response. The curve for ear No. 545 L represents a minor injury, that for 550 R a more drastic one, and that for 551 L the kind more usually found. Typically, the curve of loss showed a definite slope, with increasing impairments disclosed at the higher frequencies.

When all the ears are considered together and an average loss is computed for each test frequency, the picture is one of only moderate selectivity of impairment, with the maximum loss occurring around 5000 cycles. As shown in Fig. 2, the curve inclines from an average loss of about 5 db at 100 cycles to one of 20 to 25 db at the upper end.

This function stands in contrast with the ones for the three lower tones previously determined and reproduced here for comparison. The other functions show only a general loss, little related to frequency: the sensitivity

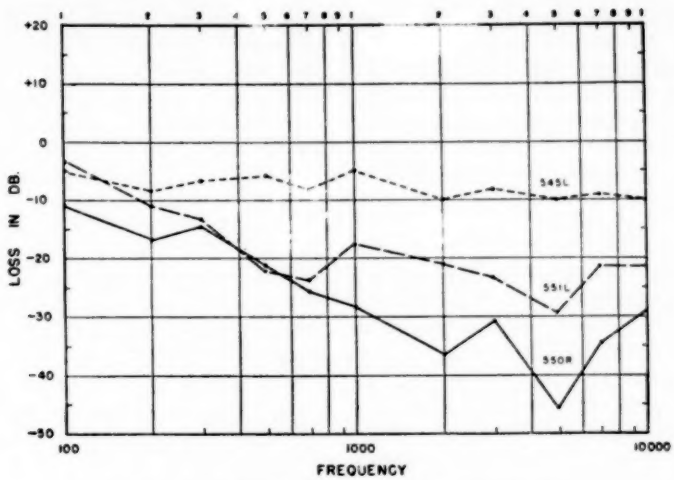


FIG. 1. Injury resulting from exposure to 10,000 cycles

to low tones and high tones alike is impaired to the extent of about 25 db. On the other hand, as the figure clearly shows, the exposure to 10,000 cycles did not so seriously affect the responses to the lower tones.

The kind of difference shown between the 10,000 cycle function and the others is found also, though only to a minor extent, among the other functions. If we examine the curves carefully we see that in general the largest losses occur for 300 cycles, and those for 1000 and 5000 cycles come in order between. It appears therefore that the injury caused by a

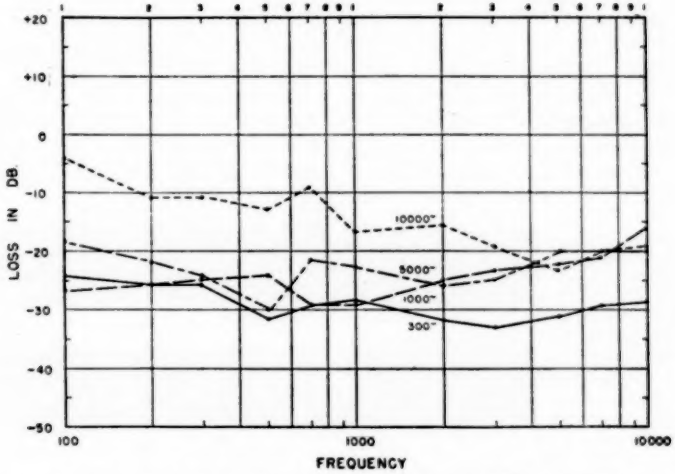


FIG. 2. Average results of injury for present group of ears and for three previous groups

tone of a given pressure and duration is less the higher its frequency. It does not appear to be a simple function of the frequency, however, and further studies are needed fully to disclose its form. This problem becomes clearer in viewing the histological results.

Histological Study

It was possible to preserve for histological study 9 of the 10 ears tested. Of these 9 ears, 8 presented a pattern of histological normality. The remaining ear (No. 509 L) showed typical exposure injury: a lesion of the organ of Corti and an associated deficiency in the tympanic lamella (the thin layer of cells adherent to the lower, tympanic surface of the basilar membrane).

The graphs of Fig. 3 portray this injury. The upper of the two graphs represents the histological condition of the organ of Corti, the full height of the ordinate denoting complete normality and lesser heights proportionately lower degrees of normality. The lower graph represents the residual proportion of tympanic lamella cells, the full height of the ordinate denoting 100 percent residual and lesser heights proportionate amounts. In each graph the abscissa represents distance from the basal end of the basilar membrane in mm.; its length represents the actual length of the membrane. The lesion indicated is sharply localized at a point about five mm. from the base or 37 percent of the way along the course of the basilar membrane. Although small, the lesion is obvious both in the organ or Corti and in the tympanic lamella.

The ear shown in Fig. 3 suffered an average loss of cochlear potentials of 22.8 db. The pattern of impairment was typical: a general slope from a negligible change in low-frequency sensitivity to a loss of about 40 db at 10,000 cycles.

One case of visible injury hardly constitutes an adequate sample. It is interesting, however, that the present histological findings are a logical extrapolation of the earlier results. It was previously found that as the frequency of the exposure tone was increased fewer and fewer ears showed histological damage. It is therefore not surprising now that injury should be rare at 10,000 cycles. Not only this, but the higher frequencies were found to produce smaller degrees of injury; and a minor lesion at 10,000 cycles is again consistent with the earlier results. It seems to be true that, with an exposure tone of constant sound pressure, the magnitude of visible injury is inversely related to the frequency of the exposure tone. As the latter increases fewer and fewer injuries become perceptible, and those that do tend to be smaller.

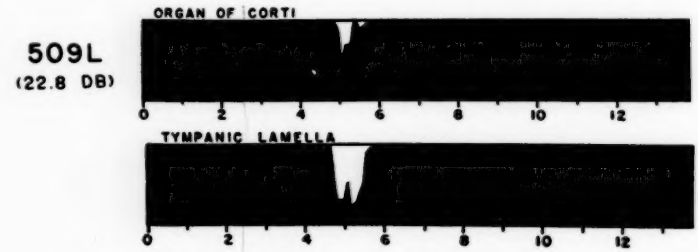


FIG. 3. Lesions produced by 10,000 cycles. Parenthetical notation gives average loss in sensitivity for the ear.

The matter of the locus of the injury cannot be emphasized because only a single case was observed, but this one observation is consistent with the others made for lower exposure tones. The region of damage, of about 37 percent of the distance up the cochlea, is nearer the basal end than the regions for the other tones. They had their centers around 51 percent for 5000 cycles, 65 percent for 1000 cycles, and 80 percent for 300 cycles. When these loci are plotted against the logarithm of the frequency they show a reasonably uniform progression.

The hypothesis advanced earlier that it is the amplitude of the cochlear vibration that measures its injurious character is further supported by these results. As already pointed out, when sound pressure is held constant, the amplitude of the vibration in any given medium varies inversely as the frequency. The amplitude of the 300 cycle tone thus is 33 times that of the 10,000 cycle tone. So we explain, in part at least, the frequency relation shown.

No doubt this relation is complicated by other variables. One is the efficiency with which the mechanical portion of the ear accepts these stimuli and transmits them inward. In the earlier studies where 300, 1000, and 5000 cycles were used an effort was made to hold this factor constant, or at least so far as its effects are disclosed in the cochlear potentials; and these tones were chosen because the sensitivity to them as seen in these potentials is practically equal for the guinea pig ear. Such is not the case for 10,000 cycles; for this tone the sensitivity is less. Still another variable that we suspect as playing a part in this situation is a regional one, representing the mobility and fragility of different portions of the sensory structures along the basilar membrane.

The lack of any strict correspondence between electrical losses and histological changes is as striking here as in the previous studies. Again, it seems reasonable to suppose that extensive sub-microscopical damage is caused by the sound. Indeed, there may be two kinds of damage, one an impairment of the capacity of the sensory cells to generate cochlear potentials, and the other a mechanical and microscopically obvious disruption of the structure.

Summary

This study extends earlier ones on the effects of over-stimulation on the guinea pig ear, by the use of a 10,000 cycle exposure tone of a standard intensity and duration. The results are consistent with the others in revealing an impairment of cochlear potentials and, in this case, minor histological damage as a result of the exposure. The injurious effect of this high tone is significantly less over the low tone range than that of other tones of like pressure and duration; in fact its effect is progressive with frequency.

These results support the earlier supposition that it is the amplitude of a tone that determines its injurious character.

(Manuscript received May 18, 1948)

References

1. Wever, E. G., & Smith, K. R. The problem of stimulation deafness. I. Cochlear impairment as a function of tonal frequency. *J. exp. Psychol.*, 1944, 34, 239–245.
2. Smith, K. R. The problem of stimulation deafness. II. Histological changes in the cochlea as a function of tonal frequency. *J. exp. Psychol.*, 1947, 37, 304–317.
3. Guild, S. R. A graphic reconstruction method for the study of the organ of Corti. *Anat. Rec.*, 1921, 22, 141–157.

THE PROJECTIVE EXPRESSION OF NEEDS. IV. THE EFFECT OF THE NEED FOR ACHIEVEMENT ON THEMATIC APPERCEPTION [1]

BY DAVID C. McCLELLAND

Wesleyan University

RUSSELL A. CLARK, THORNTON B. ROBY

Yale University

AND

JOHN W. ATKINSON

University of Michigan

Previous experiments in this series have been concerned with establishing principles for the interpretation of projective behavior. The method has been to note changes in perception (13) and apperception (2) resulting from different intensities of the hunger drive. A number of shifts in perception and in the thematic content of stories have been established which provide important clues for the detection of the strength of the hunger drive from projective records.

But the crucial experiment in the series remains to be performed. No one is particularly interested in diagnosing hunger from projective responses. The point is, do the same kinds of shifts occur for an experimentally controlled psychogenic need, or are the clues which have been discovered applicable only to some simple physiological tension like hunger?

The present experiment was designed to answer this crucial question. It was decided to choose a psychogenic need which could be aroused experimentally and to see whether it produced perceptive and apperceptive changes similar to those already noted for hunger. The need chosen was 'need achievement' or 'need mastery,' the need which *presumably* is aroused by experimentally inducing ego-involvement, according to a technique which by now is fairly well standardized among psychologists experimenting in the field of personality (1, 16, 17). The word 'presumably' is used advisedly. No one knows for certain that there is a unitary n Achievement[2] which can be satisfied by success and aroused by failure in the same way that hunger is satisfied by food and aroused by deprivation of food. However, if manipulation of the conditions of ego-involvement produces the same kinds of effects on projection as manipulation of hours of food deprivation, there will be some basis for considering the psychogenic state aroused as a need, at least to the extent that it functions like a physiological one. It was to establish this kind of parallelism of function that work began in this series with a simple physiological tension which nearly everyone would accept as a need or drive. Consequently, if the results in this experiment are in substantial agreement with those ob-

[1] This project was carried out at Wesleyan University and was made possible by a grant from the Office of Naval Research for which the authors are very grateful.

[2] The convention adopted by Murray (15) of shortening need to n will be followed throughout this paper.

tained in earlier ones in the series, it will provide evidence for the existence of higher order psychogenic needs which at least function like those at a simpler physiological level.

One of the crucial problems in this type of experiment is to find a scoring system for thematic stories which is objective enough to provide high observer agreement and sensitive enough to reflect changes in motivational states. So a further purpose of this experiment is to develop the scoring system further which was found useful for hunger (2) and to test its applicability to a more complex psychogenic need. The standardization of an objective scoring system for projective records should ultimately make possible some general principles for interpreting them. What is even more important, it should open up for experimentation the whole field of imagination which has been more or less neglected, except by the clinicians, since introspection was discredited as a fruitful approach to arriving at psychological principles.

Procedure

The materials used in the experiment consisted of some simple paper and pencil tests and some slides for thematic apperception. There were seven short tests: anagrams (4 min.), scrambled words I (3 min.), scrambled words II (4 min.), and four motor perseveration tests (4) in each of which the subject performed a writing task as often as possible in the normal manner for one min. and then backwards or in some unusual manner for one min. The total time taken including pauses between tests for instructions was about 25 min. The tests were chosen on the basis of past experience (4, 5) for a factor analysis, in connection with which they will be described in detail (6). Their chief function here was to provide the basis for inducing ego-involvement. That is, there were two main fundamentally different conditions under which the tests were administered. In one (hereafter referred to as *relaxed*), the test administrators were introduced by the instructor at a regular class session as some graduate students who were trying out some tests. This orientation was reinforced by further remarks by the 'graduate student' to the effect that these tests had been recently devised, were still in the developmental stage, and that data were being collected in order to perfect them. Throughout, the emphasis was clearly on the fact that the experimenters were interested in testing the tests and not the students. These instructions were designed to create an easy relaxed atmosphere in which the need for achievement was at a minimum.

In the other main condition (hereafter referred to as *failure*), the administration of the preliminary tests was quite different. After the experimenters had been introduced to the class by the instructors they began passing out the test booklets with no explanation as to the purpose of the experiment. The only remarks made dealt with the first test—anagrams—and the necessity of paying close attention to directions as the tests were timed. After completing the first test, the subjects calculated and recorded their scores on it. Then they filled out a short questionnaire which asked for: name, high school and college attended with estimated class standing in each, IQ (if known), and an estimate of their general intelligence (above average, average or below average). The purpose of the questionnaire was only incidentally to obtain information. It was primarily to get a subject ego-involved in the situation by making his test score known to himself and outsiders in relation to a lot of other achievement-related facts about him.

This aim was further supported by the following remarks then made by one of the experimenters (RAC) given from memory so as to give the impression of spontaneity:

> The tests which you are taking directly indicate a person's general level of intelligence. These tests have been taken from a group of tests which were used to select people of high administrative capacity for positions in Washington during the past war. Thus in addition to general intelligence, they bring out an individual's capacity to organize material, his ability to evaluate crucial situations quickly and accurately; in short, these tests demonstrate whether or not a person is suited to be a leader.
>
> The present research is being conducted for the Navy to determine which educational institutions turn out the highest percentage of students with the administrative qualifications shown by superior scores on these tests. For example, it has been found that Wesleyan University excels in this respect. You are being allowed to calculate your own scores, so that you may determine how well you do in comparison with Wesleyan students.

At this point the experimenter quoted norms for Test 1 that were so high that practically everyone in the class failed and placed in the lowest quarter of the Wesleyan group. It was then explained that Test 1 was the single most diagnostic test in the battery and thus an individual's standing on this first test would be a good indication of how well he might expect to do on the test as a whole. The rationale of giving these instructions after the first test rather than before was partly to place their first scores near the actual comparison with the norms and partly to provide a basis for testing the initial comparability of the groups in another part of the experiment (6).[3]

After this the subjects went on and completed all the paper and pencil tests. At the end they added up individual test scores to obtain a total score and again were given falsely high norms "so that they could see how well they had done as a whole in comparison with Wesleyan students."

That these instructions succeeded in producing very different effects on the students was obvious to even the most casual observer. Under the ego-involving instructions, they worked hard and quietly, and there were various obvious expressions of dismay when the norms were announced. There was no indication that the instructions were disbelieved. Under relaxed conditions the subjects were as a whole more relaxed and gave the impression of enjoying the tasks as they would a series of parlor games.

At the conclusion of the paper and pencil tests used to arouse different need states the other experimenter (JWA) read the following instructions:

> This next test is a test of your creative imagination. A number of pictures will be projected on the screen before you. You will have 20 seconds to look at the picture and then five minutes to make up a story about it. Notice that there is one page for each picture. The same four questions are asked. They will guide your thinking and enable you to cover all the elements of a plot in the time allotted. Plan to spend about a minute on each question. I will keep time and tell you when it is about time to go on to the next question for each story. You will have a little time to finish your story before the next picture is shown.
>
> Obviously there are no right or wrong answers, so you may feel free to make up any kind of a story about the pictures that you choose. Try to make them vivid and dramatic, for this is a test of creative imagination. Do not merely describe the picture you see. Tell a story about it. Work as fast as you can in order to finish in time. Make them interesting. Are there any questions? If you need more space for any question use the reverse side.

In addition to the two major conditions of administration so far described, four other attempts were made to arouse a third intensity of n Achievement which would provide the three points desirable for establishing trends and for making the data comparable with the hunger experiment (2). In the first place a *success* group was created by announcing norms after the first test and at the end which were so low that all or nearly all the students succeeded as compared with Wesleyan students. This was supposed to satiate n Achievement; but a preliminary analysis of the results indicated that while this seemed to be partly true, the need aroused by the ego-involving instructions persisted into the subsequent thematic apperception test which was interpreted as a further test of ability. Consequently, the position of this group on the need continuum was not clear and the results from it did not seem worth reporting here. Secondly, there was a simple *ego-involved* group in which no norms were announced. It was expected that these *S*s would reflect an aroused n Achievement which would be purer and not contaminated by recent experiences of failure or success. However, the stories written under this condition were so tense, inhibited, and cautious (cf. 14) that it proved difficult to analyze them, and the results from this group will also not be reported here.

The results from the final two variants on the major conditions proved meaningful and will be reported. In the first of these a group of Wesleyan students was run in a *neutral* but not 'relaxed' atmosphere. That is, they were task-oriented rather than ego-oriented (cf. 1), but they were asked to cooperate seriously and to work hard on the tasks so that adequate norms for them could be established. The reason for these instructions was to get a somewhat higher n Achievement tension than under the relaxed

[3] It goes without saying that the instructions contained nothing that was true, or for that matter nothing that was completely false. The references to the Navy and to the Washington administrators (cf. 18) were all partly true and could be checked by over-curious psychology students. Every other precaution was taken to prevent leaks about these instructions since they were crucial to the whole experiment. Different classes at the same institution were run on the same day and the procedure was not 'exposed' to the students at the end or even to the instructors (whose kindness in cooperating under the circumstances is greatly appreciated).

condition in order to maximize individual differences as part of another experiment (6). In the final group an attempt was made to get an intenser n Achievement aroused, by giving the *Ss* a taste of success by quoting low norms after the first test followed by an even greater failure at the end induced by quoting high norms. This will be referred to as the *success-failure* group.

There was no indication in the ego-involved groups that the projective tests were not still part of the program of testing for administrative ability. The slides used for eliciting the written stories consisted of two especially chosen for this experiment (two men in overalls looking or working at a machine; a young man looking into space seated before an open book), followed by two taken from the Murray Thematic Apperception Test ('father' talking to 'son'—TAT 7 BM; boy and surgical operation—TAT 8 BM). The pictures were chosen to suggest achievement—either at a specific task or general level and in school-related and unrelated situations.

The *Ss* were all male, a majority veterans, and all college students taking various psychology courses at the University of Connecticut (at Storrs and the Ft. Trumbull extension); New Britain State Teachers College, Trinity College, and Wesleyan University. They were run in regular class room periods either in the summer of 1947 or the Spring of 1948. The entire testing time which included some tests of perceptual inference reported elsewhere (14) took from 70–80 min., except for the *success-failure* condition in which it was necessary to cut out the last three of the motor perseveration tests to finish within a normal 50-min. class period.

Scoring.—The stories were scored according to the same general system used in the hunger experiment (2) with additions and modifications necessitated by the more complex nature of the need involved. Detailed scoring criteria cannot be given here for lack of space, but they have been reported in full elsewhere (12). The following brief descriptions will serve at least to identify the major categories used.

g, t, or u I: Achievement imagery is scored either general (g I) or task (t I); stories with no achievement imagery are scored as unrelated (u I). To be general, achievement imagery must deal with some long term problem of getting ahead at the ego ideal level (career, schooling, inventing something, etc.). Everything else, particularly the specific task situation, was classified as t I.

Ach or D th: Themas or plots are scored if the achievement imagery is central to the story. If the plot is concerned with someone who is in an achievement difficulty which has or is anticipated as having serious long term effects, it is scored as deprivation thema (D th); otherwise it is an achievement thema (Ach th), though there may be many difficulties in the way of the goal.

d p or w: Deprivations or blocks in the path of progress or indications of past failures, i.e., things not running smoothly. If the trouble is with the person himself it is scored d p; if it is with the world it is scored d w. D th was not scored for d also unless there was some secondary and separate source of hindrance.

N p or g: Need for achievement is stated in the story either at the personal level ("He wants to be a doctor") or at the general level ("He wants to serve humanity").

I +, −, or o: Instrumental activity which is either successful (I +) unsuccessful (I −) or of doubtful outcome (I o). The person in the story must do something (even if only think or decide) about achieving his goal which is separate from the statement of the situation and the statement of outcome: e.g., "the boy graduates from school" is scored for outcome but considered too passive to represent instrumental activity.

Ga +, −, or o: Anticipations of outcomes (goal responses) which may be either of success (Ga +, "He is thinking of the day when he'll be famous") or failure (Ga −, "He is worred about what will happen") or neither (Ga o, "He is wondering what will happen").

nu or ho P: Nurturant or hostile press. Some person in the story is either actively helping or hindering the person working for achievement. The hindrance must be more hostile than a static block (see d w above).

S: Substitution. A person who meets with an obstacle in his achievement instigation-action sequence adopts a substitute instrumental act or substitute goal response ("He drowns his sorrows in a tavern").

G or G' + or −: Goal responses which occur either within the story (G) or at the end (G') and which may be either positive affect ("He was happy in his new job") or negative affect ("The boy is worried over having flunked his exam").

O +, −, or o: Outcomes of the whole story are judged according to whether they are happy (O +), unhappy (O −), or doubtful (O o). Finer breakdowns were made but did not prove useful. The total outcome was not necessarily the same as that for the instrumental activity and was also separate from the final affect (G'). For example, "They fixed the machine" is scored O + but not G' +, because it doesn't say they were pleased about it.

The following story illustrates how the scoring was used. After each word or phrase scored is written in parenthesis the scoring symbol applicable:

1. What is happening? Who are the persons? —"The boy is being talked to by his father,

maybe something about what has happened in school, or he may be planning to get married."

2. What has led up to the situation—that is, what has happened in the past?—"He may have flunked out of school (D th) and is being lectured on what is expected of him (nu P)."

3. What is being thought—what is wanted? By whom?—"The father wants the boy to make good (N p), he is thinking that he wants his son to follow in his footsteps and make good in life (Ga +, g I)."

4. What will happen? What will be done?—"The boy will do his best, will go back to school because he has learned a lesson he will never forget (I +). He will make good this time and be a success (O +)."

As this example shows, the scoring was not done from the viewpoint of a single character with whom the writer supposedly identified, although this is the usual method of procedure. Thus the father's wish for the son's success is scored N p (father's viewpoint) and the father's help is scored nu P (son's viewpoint). The rationale for this was the conviction that determination of the person with whom the writer identified was often difficult and that it was not necessary—e.g., in this instance it is just as possible to suppose that the writer is projecting his wish to do well into the father figure as into the son. Note also that the second statement of the father's wish is not scored again. A given category is scored only once per story no matter how many times it appears.

This example shows how decisions on the scoring of a specific item were affected by the total context and by the scoring of other items. Thus, it is not until the whole story is read that the thema is clearly one of achievement and not marriage, and it is not until the third paragraph that the decision can be made that the imagery is general and the thema one of deprivation. The factors which lead to this decision are the presence of Ga + and N p (see above definitions). It was recognized that the interdependence of scoring categories was not wholly desirable from the statistical viewpoint, but it soon became obvious that the interdependence existed at the intuitive level for categories like g I and D th, no matter how discretely the definitions might be drawn. Hence, it seemed best to state as

TABLE I

THE NUMBER AND PERCENTAGE OF ACHIEVEMENT-RELATED STORIES WRITTEN UNDER RELAXED, FAILURE, AND SUCCESS-FAILURE CONDITIONS
The number of stories in each condition is 156
(39 *Ss* × 4 stories)

	Relaxed		Failure		Chi-square	Success-Failure	
	N	%	*N*	%		*N*	%
Imagery							
task t I	73	46.8	56	35.9	3.82	47	30.1
general g I	26	16.7	75	48.1	35.16	85	54.5
unrelated u I	57	36.5	25	16.0	16.94	24	15.4

Chi-square is 3.84 and 6.64 at the .05 and .01 levels of significance, respectively.

TABLE II

THE NUMBER AND PERCENTAGE OF THE ACHIEVEMENT-RELATED STORIES WRITTEN UNDER DIFFERENT CONDITIONS SHOWING VARIOUS STORY CHARACTERISTICS RELATED TO THE DESCRIPTION OF THE SITUATION

Number of Stories:	Relaxed 99		Failure 131		Chi-square	Success-Failure 132	
	N	%	*N*	%		*N*	%
Plot Ach th	59	59.6	83	63.4		98	74.2
D th	6	6.1	25	19.1	8.11	14	10.6
Obstacles d p	12	12.1	23	17.6		24	18.2
d w	22	22.2	17	13.0	3.41	21	15.9

Chi-square is 3.84 and 6.64 at the .05 and .01 levels of significance, respectively.

TABLE III

THE NUMBER AND PERCENTAGE OF THE ACHIEVEMENT-RELATED STORIES WRITTEN UNDER DIFFERENT CONDITIONS SHOWING VARIOUS STORY CHARACTERISTICS RELATED TO THE CHARACTERS' REACTION TO THE SITUATION

Number of Stories:	Relaxed 99		Failure 131		Chi-square	Success-Failure 132	
	N	%	*N*	%		*N*	%
Need stated: N p&/or g	21	21.2	58	44.3	13.30	64	48.5
Instrumental acts I +	9	9.1	41	31.3	16.29	31	23.5
I −	7	7.1	3	2.3		6	4.5
I o	6	6.1	10	7.6		7	5.3
Anticipatory goal response Ga + &/or Ga −	15	15.2	47	35.9	12.33	60	45.5
Ga o	6	6.1	11	8.4		15	11.4

Chi-square is 3.84 and 6.64 at the .05 and .01 levels of significance, respectively.

TABLE IV

THE NUMBER AND PERCENTAGE OF THE ACHIEVEMENT-RELATED STORIES WRITTEN UNDER DIFFERENT CONDITIONS SHOWING VARIOUS STORY CHARACTERISTICS RELATED TO THE OUTCOME OF THE SITUATION

Number of Stories:	Relaxed 99		Failure 131		Chi-square	Success-Failure 132	
	N	%	*N*	%		*N*	%
Press: ho P &/or nu P	10	10.1	29	22.1	5.82	20	15.2
Substitution: S	2	2.0	11	8.4	3.20*	7	5.3
Goal response: G &/or G′ +	4	4.0	33	25.2	18.61	31	23.5
G &/or G′ −	18	18.2	34	26.0		32	24.2
Outcomes: o +	47	47.5	63	48.1		69	52.3
o −	16	16.2	12	9.2		24	18.2
o o	36	36.2	56	42.8		39	29.5

* Corrected for continuity.
Chi-square is 3.84 and 6.64 at the .05 and .01 levels of significance, respectively.

explicitly as possible any other categories that were usually taken into account in the normal process of arriving at a judgment on a given category.

The way in which all the stories were scored will be described in full under results in the section on reliability of the scoring. It involved two judges working together without knowledge of which of three groups (neutral, failure, and success-failure) the stories belonged to.

RESULTS

The main results of the experiment are shown in Tables I, II, III, and IV, which summarize the frequency of appearance of various scoring categories for the relaxed, failure, and success-failure conditions. The results from the neutral condition, which generally fell between the relaxed and failure conditions, will be reported only in summary form. In all the tables increases or decreases from the relaxed to the failure condition are tested for significance by means of chi-square.[4]

[4] It should be noted that the population considered here is number of stories, rather than, as

Table I shows that there is a large and very significant increase in the number of stories dealing with general or long term achievement while there is a decrease in the number of stories with no achievement imagery and of those with task achievement imagery.

Because of this shift the method of computing percentages in Tables II, III, and IV has been changed. Since there were significantly more stories with achievement imagery in the two failure conditions, there is a correspondingly greater opportunity for other achievement-related characteristics to appear in these conditions. But the important question is: given an achievement story, are there significant differences in its internal characteristics when written under different conditions? Hence, the results for further characteristics are presented as percentages not of all stories but only of the achievement stories in each condition.

Table II shows only a significant increase in the number of deprivation themas. A comparison of the failure with the success-failure results suggests that the former reflect sensitively the *S*s' greater failure experience since the combined thema totals are nearly identical for the two conditions (82.5 and 84.8 percent respectively), both being considerably larger than the same figure for the relaxed condition (65.7 percent).

Table III shows a larger number of significant shifts in the forward-looking, striving aspects of the stories. This table indicates that an aroused n Achievement increases the likelihood that characters in the story will be described as wanting to get ahead (N p), as doing something successful about getting ahead (I +), and as thinking in advance about success or failure (Ga + or Ga −). In Table IV the shifts appear in the number of people seen as aiding or hindering achievement (nu or ho P) and in the frequency with which positive affect is specifically mentioned, either in the course of the story or at the end (G or G′ +). It is interesting to note that there are no significant changes in the outcome category, despite the fact that most of the present *a priori* systems for scoring the TAT (3, 15, 20) lay emphasis on this characteristic, and despite the fact that a far more elaborate breakdown of different types of endings was actually made than is reported here. There is one shift in outcomes, however, for the success-failure group, which gives significantly fewer doubtful outcomes to its stories than does the failure group. This suggests that repeated failure may cause an unwillingness to state the outcome of an achievement sequence, especially an unfavorable (o −) outcome, since this is the category which is reduced in the failure group.

A tabulation of the frequency of appearance of each story characteristic for each *S* was made as a basis for obtaining a single summary n Achievement score. The characteristics which showed a significant increase in Tables I–IV from the relaxed to the failure condition were scored +1 and those which decreased were scored

in the hunger experiment, the number of *S*s showing a characteristic at least once. The latter measure, while easier to interpret statistically, is not as applicable to the data of this experiment because of the much greater frequencies obtained for many of the need-related categories. The authors realize that a chi-square test of significance applied to repeated measures from the same *S*s is hard to interpret because of the peculiar nature of the universe to which the inference is made, but have decided to use it for two reasons: (1) other statistics appear to have even more serious objections, and (2) the differences found to be significant for all four pictures and used in calculating the final n Achievement score were also significant when only the results from the single most diagnostic picture (TAT 7 BM) were used.

−1.[5] Thus, there were seven positive characteristics (g I, D th, N, I+, Ga + or −, nu or ho P, and G or G′ +) and two negative characteristics (t I and u I). The results from the success-failure group were not taken into account in developing the scoring system because it was felt that the need state aroused might be more complex than in the straight failure group.

Table V presents the mean n Achievement scores for each condition. The means for the relaxed and failure conditions differ very significantly as they should from the way the scoring system was devised. The success-failure condition was almost exactly equal to the failure condition as had been indicated by the comparisons in Tables I–IV. The neutral condition showed a moderate need strength, by this scoring system, which was significantly greater than the relaxed condition and significantly less than the two failure conditions. This last comparison is particularly important methodologically, because the papers from these three groups had been mixed together and were all scored together without the judges' knowing to which group any paper belonged. Thus, with all possibility of bias removed, there is still a significant mean difference in n Achievement score between a presumed low and high intensity of induced n Achievement.

TABLE V

MEAN n ACHIEVEMENT SCORES FOR THE RELAXED, NEUTRAL, FAILURE, AND SUCCESS-FAILURE CONDITIONS

N	Relaxed 39	Neutral 39	Failure 39	Success-Failure 39
Mean	−1.00	3.13	5.82	6.00
σ_m	.46	.69	.82	.73
Diff.		4.13	2.69	.18
$\sigma_{diff.}$		.83	1.07	1.14
Critical ratio		4.98	2.51	.16
P		<.01	<.02	>.50

Reliability of the scoring.—A matter of considerable methodological importance, in view of the present tendency of experimentalists to eschew free verbal reports, is the reliability of the scoring system used here. Consequently, reliability was studied intensively from three different angles. First, an attempt was made to determine the extent to which the judges agreed on a given category for a particular story. Since agreement is almost certainly a function of amount of the judges' experience with the scoring system, a measure of it was taken at the end of the scoring, after one of the judges had had a year's previous experience with the system amounting to the scoring of at least 3000 stories and the other had scored at least 1000 stories. The two judges always worked together, one reading the story aloud so that both could form tentative judgments independently, which were discussed, if they differed, in making the final decision. At the time the test was made they were spending on the average of two to three min. per story, or at the most from five to ten min. per *S*. The test consisted of drawing 10 records at random from the neutral, failure, and success-failure groups and rescoring them. The index of agreement was computed by dividing twice the agreements by the sum of the items scored on each of the two occassions. It turned out to be 291/321, or 91 percent.

[5] A scoring system was also tried which weighted the changes in accordance with the improbability of their having occurred by chance. However, the correlation between weighted and unweighted scores for the 39 *S*s in the neutral condition was .956, confirming the high relationship found in other studies and leading to the adoption of the simpler unweighted or unit scoring.

Secondly, reliability was approached from a less conservative viewpoint by attempting to measure the extent to which the totals are stable for a given category on two judgments of the same records. This is more to the point in estimating the dependability of group shifts, since judges may quite possibly miss a catefory in one story and pick it up somewhere else leaving the total the same though not the percentage agreement. To check this, the stories written by the 39 relaxed *Ss* were completely rescored after all the other stories had been finished.[6] There was a nearly significant increase in the proportion of stories scored as containing achievement imagery, due to a conscious liberalization in the judges' set, but this increase did not change the ratio of general to task imagery or any of the other categories scored. Seventeen out of 22 of the category totals were within three points of each other.

Also relevant to this point is the comparability of the totals for various categories for the failure and success-failure groups. They are very close in nearly every case in Tables I–IV, and the mean overall n Achievement scores are practically identical. This shows that category totals are apt to be quite stable even when obtained on two different groups of *Ss*, and even when there are minor differences in the method of arousing n Achievement.

In the third place, the reliability of an individual's overall n Achievement score was tested by correlating the scores obtained for 30 individuals on two different scoring occasions. The product moment correlation was .946, indicating fairly high stability of an individual's score for his whole record. Furthermore, this correlation is probably conservative, since 20 of the 30 *Ss* came from the relaxed group, which reduced the range of scores, and since the scoring was done much more hastily on both occasions than it normally would be in a clinical situation.

[6] An attempt was made to mix the relaxed stories with others in the rescoring but it was not continued beyond the first 10, since the judges who had scored these same stories several times before easily recognized them as being very different from the others with which they were mixed. Any further attempt to conceal their identity seemed a waste of time.

Discussion

Validity of n Achievement score.—No one can deny that there are differences in the story characteristics which appear in the relaxed as compared with the failure condition, but is it proper to assume that these differences represent a difference in the need for achievement in the two groups? This is the central problem of validity, of whether the score derived from these differences measures anything of importance, or more particularly of whether it measures the n Achievement which it is supposed to measure. There are two kinds of evidence which argue that it is a valid measure of n Achievement.

The nature of the procedure used to arouse the need provides the first basis for assuming that n Achievement was more intense in the two failure conditions. In discussing what we have labelled n Achievement, after Murray (15, p. 164), Sears states: "There are many names for this learned drive: pride, craving for superiority, ego-impulse, self-esteem, self-approval, self-assertion, but these terms represent different emphases or different terminological systems, not fundamentally different concepts. Common to all is the notion that the feeling of success depends on the gratification of this drive, and failure results from its frustration" (17, p.

236). This suggests that the experimental operations which will satiate and arouse the drive are success and failure. However, the success and failure must be in relation to some achievement goal which the *Ss* have set for themselves. That is, in the case of a physiological need like hunger, it is only necessary to deprive the *Ss* of food to arouse the drive, since the organism automatically by the consumption of energy produces in time a need for food. But in the case of a psychogenic need it is necessary first to induce *Ss* to want some goal like achievement. In the present experiment that was supposedly done by giving the *Ss* an opportunity to perform on some tests which were described to them in such a way that doing well should lead them to feel increased pride, self-esteem, self-approval, feelings of success, etc. Since these terms define what is commonly meant by the striving for success or n Achievement, if the instructions and the tests were such as to arouse these feelings, then by definition n Achievement was aroused in the failure and success-failure groups.[7] And it does seem reasonable to assume that the attainment of high intelligence and leadership as suggested in the instructions are two goals which in our society would lead to the feelings mentioned.

Granted that n Achievement was aroused by the instructions, it further seems reasonable to suppose that failure-frustration would lead to a heightened need. Although this assumption is supported by experimental evidence (8) and by the deprivation method of arousing physiological drives, we recognize that it may complicate the resulting picture here. That is, granted that failure does heighten n Achievement, it may also lead to the projection of material which is specific to the experience of failure rather than characteristic of a 'pure' heightened need.[8] It was this conviction that led to the collection of stories from an ego-involved group which had had neither success nor failure. Unfortunately, as has been reported under procedure, for reasons given fully elsewhere (cf. 14) the *Ss* in this group were too inhibited to write stores which could be readily analyzed. So the main comparisons had to be made between a relaxed condition and a condition in which n Achievement was augmented by failure.

The comparison with the effect of hunger on similar stories (2) provides the second basis for arguing that a need has been aroused by the experimental conditions. However unwise it may prove to be to have used failure to heighten the need intensity, it serves to make the need-arousal method more nearly comparable to the deprivation used to increase hunger (2). Consequently, it becomes more legitimate to ask, what is the evidence that food-deprivation and achievement-deprivation affect imagination in the same way?

Table VI provides the positive evidence that the two needs have the same general effect. The case rests

[7] Since these instructions are also the ones commonly called 'ego-involving' by other workers in the field (cf. 1), it is apparent the authors believe that ego-involvement and n Achievement arousal are the same thing under certain conditions.

[8] The fact that the failure group showed more deprivation themas than the success-failure group, while both showed about the same high number of themas as compared with the relaxed group, would support this proposition. That is, one could argue that heightened need tension results in more achievement imagery central to the plot, but that failure as a method of increasing this tension shifts some of this plot or thema imagery to the deprivation category.

TABLE VI

A Comparison of the Story Characteristics Showing Significant Changes for Both Increased n Food and Increased n Achievement

1. An increase in the number of plots dealing primarily with deprivation of the goal in question (D th)
2. An increase in the number of times that characters in the stories were said to want or wish for the goal in question (N)
3. An increase in the mention of instrumental activities which are successful in dealing with the need-related problem (I+)
4. No change in the number of plots dealing with direct attainment of the goal (F th or Ach th)
5. No change in the amount of substitute activity, in instrumental activity of unsuccessful or doubtful outcome, or in negative affect (represented by subjective hostility in the food experiment)

largely on the first three items (D th, N and I +), since failure of categories to shift may mean failure of the scoring system at some point. Even so the evidence is impressive, considering the fact that a complex psychogenic need like that for achievement might be supposed on *a priori* grounds to differ extensively from a simple primary need like hunger.

There is also some negative evidence, i.e., instances of categories which shift in one experiment but not in the other. But these can rather easily be explained in terms of differences in procedure in the two experiments. For example, the biggest lack of correspondence in the two experiments was in the way a higher need decreased the favorable aspects in the food stories and increased them in the achievement stories. A case in point is the decrease in friendly press for hunger and the increase in nurturant press for n Achievement. This can be explained by the fact that the two control groups were not equivalent. The one-hour hunger group was *satiated* with respect to hunger, whereas the relaxed group in the present experiment could best be described as *unmotivated* with respect to n Achievement. Satiation undoubtedly carries over to increase the frequency of appearance of favorable story aspects, as has been shown for the n Achievement situation when success is given the *Ss* (**12, 14**). Since the low need groups therefore doubtless differed initially in the amount of favorable material projected, it is not surprising that high need groups in the two experiments produced different or even opposite effects. Other incongruencies between the two experiments are largely due to changes in the scoring system necessitated by the greater complexity of n Achievement (e.g., the general imagery category).

If one notes the major agreements and explains away in this manner the disagreements, Table VI can be said to supply considerable support for the argument that the conditions of this experiment induced a state in the *Ss* which affected their imagination in the same general way as an increase in hunger. To the extent that one accepts hunger as a need, it would therefore seem valid to refer to the state induced by ego-involvement and failure as a need. Even if one grants this, however, it must of course still be shown that the situationally induced need affects apperception in the same way as a strong character need would, as clinically or otherwise defined. This ultimate problem of validity must await further study.

Clinical applications.—In the meantime, the data are sufficiently clearcut to provide some guidance for the person working with the TAT clinically. They suggest in the first place what story characteristics are apt to be important as indicators of need strength. Although the validity of these indicators is by no means finally established, they do represent an advance over the logical or *a priori* validity earlier workers in the field have been forced to assume for their

scoring systems. In the second place, the data suggest to the clinician that the conditions of administration of the TAT are of considerable importance in determining the dynamic content of the stories. Stories written under relaxed, neutral, and failure conditions differed so much in the present experiment as to suggest more caution than has heretofore been indicated in assuming that the basic personality picture given by the TAT is not influenced by recent experiences (3, 7). Our results suggest strongly that the clinician should be careful to investigate such matters as how the subject conceives of the test, his reason for taking it, his relation to the testor who may or may not have given him other tests that have involved success or failure, etc.

Nature of motivation.—One of the most important implications of this experiment is suggested by a consideration of the categories which shifted in frequency when the need was presumably aroused. Most, if not all of them, appear to have a future reference—for instance, the stated wish for achievement, successful instrumental striving, anticipatory goal responses, and positive affect at the end of the story. Two other important characteristics—the increase in general imagery and the increase in deprivation themas—also appear to refer to the future on further scrutiny, the former because it is defined as involving a person's career or life work, and the latter because it is defined as a situation in which forces are at work against a person which would make him worse off in the future. In both instances the presence of stated need or anticipatory goal response was often useful in defining the category. On the other hand categories did not change which seemed to involve more of an objective description of the situation (plots, obstacles and outcomes) without the striving or anticipatory dimension. This, taken with similar earlier evidence (14), suggest that it is one of the major characteristics of motivation —at least achievement motivation—to be anticipatory or forward looking. This might seem to be a somewhat radical departure from the usual conception of a motive as a persisting deficit stimulus, but oddly enough Hull (9), working from entirely different data, has come to much the same conclusion—namely, that fractional anticipatory goal responses are the key to understanding purposeful and motivational phenomena. In fact, one can argue that the anticipatory goal responses observed in this experiment supply a kind of direct confirmation of Hull's view which has been very difficult to obtain with animals.

Methodological considerations.—Last but not least these results have an important bearing on the experimental methodology of handling verbal material. They report a method for scoring written thematic apperception stories which is sensitive enough to distinguish between the conditions under which the stories are written, which is objective enough to yield high agreement on a repeat scoring by two trained judges working together, and which is easy enough to apply quickly to an individual record. This in itself is of considerable importance in a field in which prior scoring systems have either been so complex or so dependent on clinical insight (3, 15, 20) that they are of little use to the experimental psychologist interested in studying imaginative processes.

The potential value to psychological theory of an objective scoring method for free verbal behavior is illustrated by the fact that its application in this experiment clearly indicates that phantasy does not always serve the purpose of wish-fulfillment or substitute gratification for pleasures denied in reality, an assumption which has been rather frequently made (cf. 10, p. 93). Instead, a study of the variety of story characteristics which shifted in this experiment with an increase in need supports the parsimonious assumption that imaginative behavior is governed by the same general principles as govern any behavior. For example, a variety of experiments show the same increase in instrumental activity with increased drive at the gross motor level; others, as in the standard Pavlovian conditioning, show the same increase in anticipatory

goal responses (salivation). If one grants that the principles governing imaginative behavior are no different from those governing performance when both are analyzed according to the same categories of response, then the method used here becomes a more subtle and flexible approach to the establishment and extension of those principles than the ordinary method of studying performance. Thus, for example, it would be difficult to get a performance response which would correspond to the anticipation of deprivation which follows drive arousal at the imaginative level. One might even go so far as to suggest that by the use of this method Tolman could study much more directly the 'cognitive maps' which the behavior of his rats has led him to infer are the important intervening variables in determining behavior (19).

Summary

Over 200 male college students wrote five-min. stories in response to four slides depicting achievement-related situations under the influence of various interpretations of the meaning of the story writing and several short pencil and paper tests taken just previously. The stories were analyzed completely for 39 *Ss* from each of four conditions: (1) a *relaxed* condition, in which all the tests were interpreted as being in an experimental stage, (2) a *neutral* condition, in which the tests were described as experimental but in which the *Ss* were urged to do their best to establish some norms, (3) a *failure* condition, in which the tests were interpreted as standardized measures of intelligence and leadership and in which the *Ss* wrote their stories after failing on the paper and pencil tests, and (4) a *success-failure* condition, which was the same as the failure condition except that the *Ss* succeeded on the first part of the paper and pencil tests and then failed on the whole test. The stories from a group who wrote under ego-involving instructions but without success and failure proved too inhibited to analyze, and those from a group who succeeded throughout are not reported because the meaning of the situation to the *Ss* did not seem clear. The scoring followed in general the usual analysis of an overt behavioral sequence with adaptations from Murray. The following results were obtained:

1. The scoring method, when used by two experienced judges working together, could be quickly applied (two to four min. per story), was sensitive enough to discriminate among the stories written under different conditions even when mixed together before judging, and was objective enough to yield on rescoring a 91 percent agreement for individual categories and a rescoring reliability coefficient for the n Achievement score developed of .948.

2. On the assumption that the relaxed and failure conditions represented a low and high degree of induced need for achievement, a comparison was made of the category shifts between these two groups. The following changes occurred at least at the .05 level of significance: a *decrease* in unrelated and task achievement imagery, an *increase* in general achievement imagery, achievement-related deprivation themas, stated needs, successful instrumental acts, anticipatory goal responses, nurturant or hostile press, and positive affective states. In nearly every case the success-failure condition showed the same percentages as the failure condition providing a category total stability check.

3. A single n Achievement score was computed for each individual by summing the characteristics he showed

which increased reliably for the group and subtracting those which decreased reliably. The mean n Achievement scores computed in this way increased significantly in accordance with the presumed increase in induced need from relaxed, to neutral, to the failure conditions.

The validity of these results as true measures of n Achievement is discussed particularly as it derives from a comparison with similar trends obtained with hunger and from a consideration of the experimental operations performed on the Ss. The data are further interpreted as pointing to the dynamics of the test situation as an important determiner of TAT content, as supporting a theory of motivation based on anticipatory goal responses, and as providing a method for investigating such important theoretical constructs as 'cognitive maps' and 'anticipatory goal responses' which is more sensitive than that based on the usual inferences from performance responses.

(Manuscript received for immediate publication December 11, 1948)

References

1. Alper, T. G. Task-orientation vs. ego-orientation in learning and retention. *Amer. J. Psychol.*, 1946, **59**, 236–248.
2. Atkinson, J. W., & McClelland, D. C. The projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
3. Bellak, L. The concept of projection. *Psychiatry*, 1944, **7**, 353–370.
4. Cattell, R. B. The riddle of perseveration. I. "Creative effort" and disposition rigidity. *J. Person.*, 1946, 229–238.
5. Clark, R. A. The problem of closure in mental organization. Unpublished Honor's Thesis, Wesleyan Univ., 1947.
6. Clark, R. A., & McClelland, D. C. A factor analytic integration of apperceptive and clinical measures of n Achievement. Unpublished ONR report.
7. Coleman, W. The Thematic Apperception Test. I. Effect of recent experience. II. Some quantitative observations. *J. clin. Psychol.*, 1947, **3**, 257–264.
8. Festinger, L. Development of differential appetite in the rat. *J. exp. Psychol.*, 1943, **32**, 226–234.
9. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.
10. Langer, W. *Psychology and human living.* New York: Appelton-Century, 1941.
11. McClelland, D. C., & Apicella, F. S. Reminiscence following experimentally induced failure. *J. exp. Psychol.*, 1947, **37**, 159–169.
12. McClelland, D. C., Atkinson, J. W., & Clark, R. A. The effect of different intensities of need mastery on thematic apperception. Mimeographed ONR report, 1947.
13. McClelland, D. C., & Atkinson, J. W. The projective expression of needs. I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–232.
14. McClelland, D. C., Atkinson, J. W., & Clark, R. A. The projective expression of needs. III. The effect of ego-involvement, success and failure on perception. *J. Psychol.*, 1949, **27**, 311–330.
15. Murray, H. A., et al. *Explorations in personality.* New York: Oxford Univ. Press, 1938.
16. Nowlis, H. H. The influence of success and failure on the resumption of an interrupted task. *J. exp. Psychol.*, 1941, **23**, 204–325.
17. Sears, R. R. Success and failure: a study of motility. Chapter 13 in: *Studies in personality.* New York: McGraw-Hill, 1942.
18. Thurstone, L. L. *A factorial analysis of perception.* Chicago: Univ. Chicago Press, 1944.
19. Tolman, R. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
20. Tomkins, S. S. *The thematic apperception test.* New York: Grune and Strutton, 1947.

RELATIVE SELF-ESTIMATE AND THE LEVEL OF ASPIRATION

BY ALBERTA S. GILINSKY

Tufts College

Several investigations of the determinants of the level of aspiration have shown that knowledge of the performance of others may shift the subject's goal (1, 2, 3). Such knowledge provides a frame of reference within which the individual sets his level of aspiration. Chapman and Volkmann (1) have applied the concept of anchoring to the effective determinants of this reference frame. In their experiment, the suggestion that a superior group (of authors and literary critics) had made a certain average score on a test of literary acquaintance exerted an anchoring influence which lowered the aspiration level. The suggested score of a supposedly inferior group (WPA) exercised an anchoring effect upward on the stated level of aspiration. Knowledge of the performance of a similar group (of college students) held the mean aspiration level near the suggested score.

Chapman and Volkmann conclude that ". . . before the level of aspiration has become too dependent upon direct experience with a task, knowledge of the performance of other groups may raise or lower it. "But, these authors ask, "is the amount of raising or lowering a function of the perceived difference in ability between the individual and these other groups? Our research leaves this point in question; for while we have dealt with groups which the subjects clearly regard as different from themselves, we have not attempted to measure the degree of this difference, nor to vary it."

The present experiment is an attempt to answer this question. An attempt is made to establish the precise position of the comparison group in relation to the individual. To do this, the 'I.Q.' level of each one of a series of comparison groups is given. At the same time, the subject provides an estimate of his own I.Q. By subtracting the 'I.Q.' of the reference group from the stated individual I.Q., one may measure both the direction and the degree of the perceived difference in ability between the individual and the group. The performance score supposedly made by each group (but actually identical from one group to the next) is made known to the subject, who then states his level of aspiration for the test. The level of aspiration may be expected to vary depending upon the remoteness of the reference group. A series of such groups, whose average 'I.Q.'s' cover a wide range of values, should permit the determination of the functional relation between the perceived difference in ability and the level of aspiration.

PROCEDURE

One hundred college students who had not studied psychology and were naive with reference to the purpose of the experiment served as *S*s.

Each *S* was given an individual mimeographed form, consisting of a vocabulary test covered by an instruction sheet. The instruction sheet defined the comparison group in terms of 'I.Q.' and occupation. The statement of occupation was worded as inconspicuously as possible, thus subordinating it to the 'I.Q.' while still maintaining plausibility. Different forms were used to describe each of 10 specified 'I.Q.' levels, ranging from 80 to 170. Ten experimental groups were formed, each group receiving a different set of instructions. Three groups were told that a

group of educators with an 'I.Q.' of 170 (160, or 150) had made an average score of *18* on the test; four groups were told that college students with an 'I.Q.' of 140 (130, 120, or 110) had made an average score of *18*; and three groups were told that a group of manual laborers with an 'I.Q.' of 100 (90, or 80) made an average score of *18*. All groups were told the minimum (zero) and maximum (36) possible scores. With the appropriate instructions substituted for each experimental group, the form read as follows:

> The following test is completely anonymous. Answer the two questions on this sheet. Then proceed to the test.
>
> This test is part of a battery designed to measure I.Q. A group of educators with an I.Q. of 170 were given this test and made an average score of 18. The minimum score is 0; the maximum is 36.
>
> What score do you expect to make?
>
> What is your approximate I.Q.?

As in Chapman and Volkmann's experiment, the *S*s stated their levels of aspiration in advance of performance or acquaintance with the task. The test itself was thus of little importance. It consisted of 36 vocabulary items in multiple choice form. Each word was followed by four possible synonyms, *S* being instructed to underline the one which best defined the given word.

Results

Table I shows, for each of the experimental groups, the mean aspiration level and the mean self-estimate of I.Q. There is a striking relation between the mean level of aspiration and the 'I.Q.' of the comparison group. The level of aspiration increases, almost doubles, as the 'I.Q.' of the comparison group falls from 170 to 80.

TABLE I

Mean Levels of Aspiration and Mean Estimates of I.Q. for the Ten Experimental Groups

Group	I.Q. of Comparison Group	Mean Aspiration Level	Mean Self-Estimate of I.Q.
1	170	14.5	122.5
2	160	16.4	130.0
3	150	12.9	119.0
4	140	17.6	121.6
5	130	19.4	133.3
6	120	19.3	125.0
7	110	20.1	121.3
8	100	24.6	121.7
9	90	26.4	123.5
10	80	27.2	126.4
Mean	125	19.8	124.4

TABLE II

Relative Self-Estimate and the Level of Aspiration

Self-Estimate of I.Q. Minus Group 'I.Q.'	*N*	Mean Aspiration Level	SD of Aspiration Level
61–71	1	34.0	—
50–60	5	30.4	5.8
39–49	7	25.6	2.7
28–38	5	26.0	5.0
17–27	15	23.4	3.8
6–16	8	20.0	2.9
−5–5	17	19.6	6.1
−16–−6	4	17.2	3.3
−27–−17	15	15.2	5.6
−38–−28	5	16.2	8.2
−49–−39	10	13.8	4.8
−60–−50	7	12.9	2.2
−71–−61	1	12.0	—

On the other hand, the mean self-estimate of I.Q. does not appear, from Table I, to vary in any consistent manner either with the 'I.Q.' of the comparison group or the mean aspiration level. However, when the individual cases are taken into consideration, the absolute self-estimate of I.Q. does show a positive, though low, relation to the level of aspiration. The product-moment coefficient of correlation between these two variables is .461.

The major finding of the experiment is shown in Table II and Fig. 1. The level of aspiration is significantly related to the degree of perceived difference in ability between *S* and the comparison group—the epsilon corrected for broad categories based on individual cases is .660.[1]

[1] Preston (5) has shown that special considerations arise in the case of a correlation between a difference-score and another measure, of which the correlation just quoted is a type. The coefficient in the present instance may be a func-

Table II and Fig. 1 were constructed by subtracting from *S*'s own estimated I.Q. the 'I.Q.' of the particular group presented to that *S* for comparison. The resulting differences in I.Q. ranged from −65 to 65, and may be interpreted as the perceived difference in ability between the individual and the reference group. A positive difference indicates that *S* rates his own ability above that of the reference group; a negative difference points

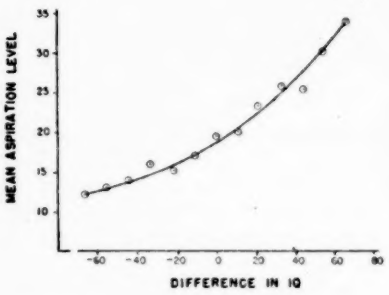


FIG. 1. Aspiration level as a function of perceived difference in ability between the individual and the comparison group

to a corresponding perceived inferiority to the group. The magnitude of the difference measures the degree of the judged difference in ability.

These differences in I.Q. were then grouped into 13 step-intervals without regard to the absolute value of either *S*'s self-estimate or the 'I.Q.' of the comparison group. The levels of aspiration corresponding to I.Q. differences falling within a given step-interval were averaged. The mean aspiration level is shown for the appropriate step-interval for each of the 13 step-intervals in Table II.

In order to ascertain the shape of the function, the mean aspiration level is plotted against the midpoint of the appropriate step-interval in Fig. 1. The midpoint is thus taken as the most representative value of a given degree of perceived difference in ability between *S* and the comparison group. As this difference increases from minus values through zero to larger and larger positive values, there is a continuous increase in the level of aspiration. The best fitting function appears to have a slight curvilinearity. In this connection it may be noted that the r for these data is .666, and the η is .694.

DISCUSSION

An analysis of the frames of reference that apparently determined the levels of aspiration shows that the suggested standard, or performance score of the comparison group, 18, provides an effective anchoring agent. The direction and extent of its influence depends upon the estimate *S* makes of his own ability in relation to the known ability of the group. If *S* judges his own ability inferior to the group, then the anchor tends to shift the aspiration level downward, below the group norm. When *S* perceives that his own ability is superior to the group, then he 'should do better' than the group average, and the anchor exerts an upward effect upon his level of aspiration.

The amount of the shift upward or downward varies with the size of the estimated difference in ability between the individual and others. Thus, the anchoring influence exerted by the standard is proportional to its psy-

tion of the absolute variabilities of estimate, and comparison group 'I.Q.'s' as well as of the three intercorrelations between self-estimate of I.Q., comparison group 'I.Q.' and aspiration level. However, using Preston's formula (Expression 4) which may be applied to the present problem, the correlation between level of aspiration and the difference between comparison group 'I.Q.' and own estimated I.Q. is found to have a similar value, namely, +.628.

chological distance from the *S*—a finding in complete agreement with the results of experiments on anchoring influences in absolute judgment. Several investigators (4) have found a functional relation between the remoteness of the anchoring value and the extension of the individual's scale of judgment. In the present study, the fact that for different *S*s the anchoring agent possesses various degrees of psychological remoteness permits the observed relation to appear. The more the *S* differs in perceived ability from the group, the greater the difference between his level of aspiration and the standard set by the group performance.

The setting of the level of aspiration thus results from the combined influence of at least two anchoring points; the known performance of another group, and the person's estimate of his own ability. In addition, anchoring effects were presumably created by the knowledge of the maximum and minimum possible scores, since no *S* aspired lower than a score of 9, and few set their goals higher than 30.

The results reported here confirm and extend the findings of Chapman and Volkmann (1). As in their study, a social determinant, knowledge of the performance of other groups, before first-hand acquaintance with the task, governs the setting of the level of aspiration. The present data also permit a definite answer to the question raised by Chapman and Volkmann. The degree to which the level of aspiration is raised or lowered is a function of the perceived difference in ability between the individual and these other groups.

Summary

This investigation concerned the functional relation between the level of aspiration and the individual's estimate of his own ability relative to a group whose performance is known. Knowledge of the average 'I.Q.,' together with the average performance score on a vocabulary test of 10 groups was funished separate groups of college students. An estimate of I.Q. and a statement of level of aspiration was obtained from each *S* before experience with the test. The difference between *S*'s stated I.Q. and the supposed 'I.Q.' of the particular group presented for comparison was taken as a measure of the perceived difference in ability.

A functional relation was observed between the two variables: the greater the perceived difference in ability between the individual and the group, the further the level of aspiration diverges from the supposed group norm.

(Manuscript received June 23, 1948)

References

1. Chapman, D. W., & Volkmann, J. A social determinant of level of aspiration. *J. abnorm. soc. Psychol.*, 1939, **34**, 225–238.
2. Gould, R., & Lewis, H. B. An experimental investigation of changes in the meaning of level of aspiration. *J. exp. Psychol.*, 1940, **27**, 422–438.
3. Hertzman, M., & Festinger, L. Shifts in explicit goals in a level of aspiration experiment. *J. exp. Psychol.*, 1940, **27**, 439–452.
4. McGarvey, H. R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, No. 281.
5. Preston, M. G. Use of the coefficient of correlation in the study of the D-score for the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 442–446.

THE EFFECTS OF PUNISHMENT ON ACTS MOTIVATED BY FEAR [1]

BY GORDON T. GWINN

Wesleyan University

Introduction

Individuals often tend to stop performing punished acts and to avoid punishing situations. In accounting for this fact some psychologists assume that punishment weakens or depresses the S-R connections involved. Others hold that it produces expectancies or negative valences resulting in avoidance of the punishing stimulus. These theoretical propositions are in essential agreement with Hilgard and Marquis' empirical principle of 'inhibition by punishment.' They hold that "*if a response is followed by a noxious stimulus it will tend to be weakened in strength and less likely to occur on the next presentation of the situation*" (1, p. 125). It should be noted that the term 'inhibition,' as used by these authors and in this paper, refers to a decrease in response *performance* rather than to any hypothetical mechanism.[2]

On the other hand, according to S-R contiguity and drive-reduction theories of learning, punishment does not necessarily have any inhibitory effects. The only learning resulting from punishment is the learning of the responses, especially escape responses, evoked by the punishing stimulus and conditioned to the cues for the punished act. Whether or not a given punishment will be inhibitory depends on whether or not these 'anticipatory' responses are incompatible and so interfere with the punished act. It also depends, according to the fear hypothesis of Miller and Dollard (3, especially p. 61), and Mowrer (4), on whether or not the dominant response to fear is incompatible with the punished act. This is because anticipatory fear-responses produce stimuli tending to elicit, innately or as a result of learning, responses which may or may not interfere with the act.[3]

It follows from these 'interference' theories that, if the response to the punishing stimulus is compatible with the punished act, *punishment will facilitate rather than inhibit an act motivated by fear.* The fear produced by the punishment will strengthen the fear-drive motivating the punished act. The act will be facilitated because response performance is a positive function of drive-strength. Furthermore, the fear-drive, and hence the facilitation, will increase with the intensity and frequency of the punishment. There will be no interference and hence no inhibition, because the dominant response to fear *is* the punished act and it is not incompatible with the response to the punishing stimulus.

[1] This report is based on a dissertation presented in 1948 to the Faculty of Yale University in partial fulfillment of the requirements for the Ph.D. degree. I am indebted to Dr. Neal E. Miller, under whose direction the research was conducted, and to the other members of the Psychology Department for many helpful criticisms and suggestions, and to my wife, Adelaide S. Gwinn, for assistance in conducting the experiments and for drawing the figures.

[2] Similarly, McGeoch (2) proposes as a valid 'empirical law' the view that punishment 'eliminates' the punished act whatever may be the underlying mechanisms.

[3] For summaries of the divergent views concerning the mechanisms mediating the effects of punishment see McGeoch (2), Postman (5), and Hilgard (6).

The experiment reported below was designed to test this facilitation hypothesis as against the view that punishment necessarily has inhibitory effects.

METHOD

Subjects.—A group of 24 experimentally naive male albino rats from three to four months old were used after three days during which they were handled in order to tame them.

Apparatus.—The circular alley diagrammed in Fig. 1 was used. Raised three feet from the floor, it was three in. wide and four in. high, and had an inside diameter of 19 in. The runway consisted of a wire grill to which alternating electric current could be delivered by means of a telegraph key. A potentiometer was used to control voltage, and a one-fourth megohm resistance in series with the grill helped to minimize the effects of fluctuations in the animals' resistance.

Four gauze-covered glass panes, individually removable from a wooden frame, constituted a one-way vision cover for the sheet-metal alley walls. Eight sections of the alley were labeled for purposes of reference.

Placed in a darkroom, the alley was illuminated through the grill by a 200-watt bulb and reflector placed 18 in. below the center. Attached under sections 6 and 7 was a black cardboard shield making the floor black, in contrast to the bright floor of the rest of the alley, and casting a sharp shadow upon the walls of the alley. Thus sections 6–7, where the punishing shock occurred during the test trials, was visually distinct from the rest of the alley. The line between the bright section 5 and the dark section 6 will be called the 'cue,' since it provided a 'danger signal' at the point where the shock area began.

The *S*s were placed in the alley through a removable hopper with flared sides placed over section 1 (replacing the glass cover for that section). A barrier attached to the hopper prevented movement from section 1 directly into section 8. A wooden block, made to fit flush with the top of the alley walls, was used to ensure the *S*s' entrance into the alley and to prevent

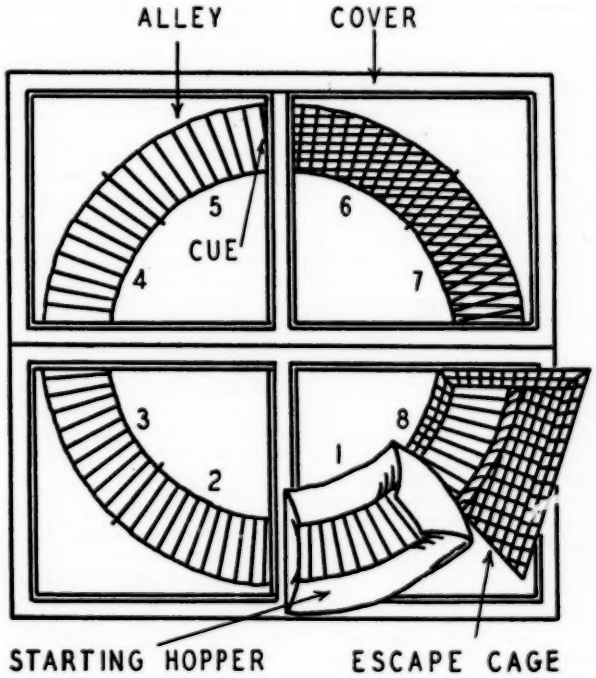


FIG. 1. Diagram of the alley

exit by way of the hopper. The only exit possible was by way of an escape cage which replaced glass over any desired section. A barrier attached to the escape cage prevented movement beyond the cage.

Two electric timers were manually operated by an assistant who also made the required recordings.

Preliminary training.—The *S*s were given preliminary training which was designed to produce a fear-drive in the alley and a tendency to run forward and escape from the alley as a consequence of this fear.

Each animal received 18 trials during which he started at section 1 and received 60-volt shocks throughout the alley until he jumped into the escape cage. To facilitate the training, the escape cage was placed over section 2 for 4 trials, then moved successively to sections 4 and 6 for 3 trials each, and finally to section 8 for 8 trials.[4] Upon escape from the alley, the animal was removed to a neutral restraining cage for 10 sec. before the next trial.

On the first few trials all the animals reacted to the shock with a variety of responses other than running forward. These responses soon disappeared, however, and rapid running forward and jumping into the escape cage occurred consistently by the end of the training session. Furthermore, on the first test trial following this training, although no shocks were received from start to 'cue' (i.e., the line between section 5 and the darkened section 6), the *S*s promptly ran that distance (the mean running time being less than two sec.). By the fear hypothesis, these facts indicate that the training sessions effectively established fear in the alley and also established running as the dominant response to fear as well as to shock. This was the response which was to be punished during the test trials.

Test trials.—One min. after completion of the preliminary training, each animal was tested under one of the following randomly selected conditions:

a. *Shock vs. non-shock.* To determine whether shock administered only as a 'punishment' for running to the 'cue' would facilitate or inhibit that response, 16 of the animals received shock only if they ran past the cue and entered the dark area of the alley (sections 6–7). The remaining eight animals received no shocks and thus provided an experimental 'control' of factors other than the punishing shock which influenced the running tendency.

[4] During the first few trials, occasionally an *S* stood motionless on his hind legs, apparently reducing the effectiveness of the shock. When this happened, the experimenter prodded the animal's feet and movement was resumed.

b. *Weak shock vs. strong shock.* To determine the effects of different intensities of shock, eight of the shocked animals received shocks of 60 volts (the same as in the preliminary training) while the remaining eight received 120-volt shocks (double the training voltage).

c. *Shock-100% vs. shock-33%.* To determine the effects of different frequencies of shock, half of each of the above shock groups received shocks on every trial; the other half received shocks on the first trial but only on every third trial thereafter. The shock-33% condition also provided a basis for comparing the effects of shock trials as against non-shock trials for the same animals.

Under all conditions the animals were started at section 1, and the escape cage remained at section 8. On all shock trials the shock went 'on' as soon as the animals passed the cue. Once the shock was on, it remained on throughout the alley until they ran into section 8, which was not electrified during test trials. It is important, for theoretical reasons to be discussed later, to note that all the animals were observed to respond to the shock by dashing forward into section 8. Since no returning to section 5 occurred, none of the animals received shocks anywhere except in the dark sections 6–7, and for all of them the response accompanied by termination of the shock was invariably the response of running forward.

Upon escape from the alley, the animals were removed to the restraining cage for an interval of 10 sec. before the next trial.[5] Trials were repeated until the running time from start to cue exceeded a 'stopping' criterion of 10 sec.

Measures.—On each trial the time required to run from start to cue, and also the time from cue to escape, were measured to the nearest tenth of a sec. These scores were converted to speed, i.e., the number of sections run per sec., and the mean of each individual's speed on all trials was computed. Then the group mean of these individual mean speeds was computed for each experimental condition. The mean number of trials required to reach the 10-sec. stopping criterion was also computed for each condition.

The speed of running and the number of trials run before reaching the stopping criterion were

[5] Occasionally *S*s of the non-shock group turned around and started to return to section 5. Such movement was blocked by the insertion of a metal gate through the grill in section 5. Also, occasionally, animals of this group remained in sections 6–8 longer than 10 sec. and, when tapping the apparatus failed to result in jumping into the escape cage, they were removed directly from the alley. This was unnecessary for shocked animals because they never returned to section 5 and always jumped into the escape cage.

used to measure any facilitation or inhibition of the punished act of running to the cue. In addition, since the effects of punishment whether facilitatory or inhibitory were expected to influence the frequency and location of any tendencies to stop running, the location of every instance of stopping was recorded.

RESULTS

Evidence of facilitation by punishment

a. *Effects of shock vs. non-shock.*— It may be seen in Table I that the animals that received the punishing shock at the cue ran to the cue faster than the animals that were not shocked. The differences between the mean speeds for all trials run by the non-shock as against the shock groups, whether the shock was strong or weak, every trial or every third trial, are statistically highly significant in all cases.

The fact that the punishing shock at the cue produced faster running to the cue than non-shock is shown graphically in Fig. 2. The curve for shocked animals in this figure also shows that the initial running tendency was increased by the punishment during the first half of the test trials. All but one of the 16 shocked animals ran faster, on at least one of the first half of trials, than they did on the initial trial. This contrasts with the curve for non-shock groups which reflects the fact that all eight of these animals tended to slow down immediately, and at no time did any of them ever run faster than on the first trial.

Animals shocked at the cue not only ran to the cue faster than those not shocked, but also they tended to perform this act for more trials. Table II shows that shocked animals ran more trials before reaching the stopping criterion than non-shocked. The differences in the mean for the non-shock group as compared with the

TABLE I

ANALYSIS OF EFFECTS OF VARYING INTENSITY AND RELATIVE FREQUENCY OF SHOCK UPON MEAN SPEED (NUMBER OF SECTIONS PER SECOND) OF RUNNING FROM START TO CUE *

A. Group Means

Test Condition	Strong Shock (SS)	Weak Shock (WS)	Non-Shock (NS)	Shock-100% (S100)	Shock-33% (S33)
Number	8	7	7	7	8
Mean	2.70	2.06	1.43	2.46	2.34

B. Differences between Means

Means Compared	Effect of Intensity			Effect of Frequency		
	SS-WS	SS-NS	WS-NS	S100-S33	S100-NS	S33-NS
Difference	.64	1.27	.63	.12	1.03	.92
Level Sig.**	1%	1%	1%	35%	1%	1%

* Analysis of the effects of varying intensity is based on the same data as those used for frequency, but the data were recombined.

One weak-shock-100% and one non-shock animal were excluded from their respective groups in computing mean speeds. This was done because each ran only one trial. A 'mean' of a single score is unreliable on a priori grounds and, furthermore, these scores deviated from their group means by 4.88 and 3.66 sigmas respectively.

** The *t*-tests of statistical significance described by Lindquist (7) were used in all the analyses in this paper. In all instances the 'levels of significance' stated refer to the percent of cases in which differences in the same direction as large or larger than those obtained would be expected to occur by chance; all percentages smaller than one are reported as one percent.

TABLE II

ANALYSIS OF VARYING INTENSITY AND RELATIVE FREQUENCY OF SHOCK UPON THE NUMBER OF TRIALS RUN BEFORE REACHING THE 10-SECOND STOPPING CRITERION *

A. Group Means

Test Condition	Strong Shock (SS)	Weak Shock (WS)	Non-Shock (NS)	Shock-33% (S33)	Shock-100% (S100)
Number	8	8	8	8	8
Mean	40.50	16.63	9.25	32.00	25.13

B. Differences between Means

Means Compared	Effect of Intensity			Effect of Frequency		
	SS-WS	SS-NS	WS-NS	S33-S100	S33-NS	S100-NS
Difference	23.87	31.25	7.38	6.87	22.75	15.88
Level Sig.	2%	1%	10%	30%	1%	5%

* Analysis of the effects of varying intensity is based on the same data as those used for frequency, but the data were recombined.

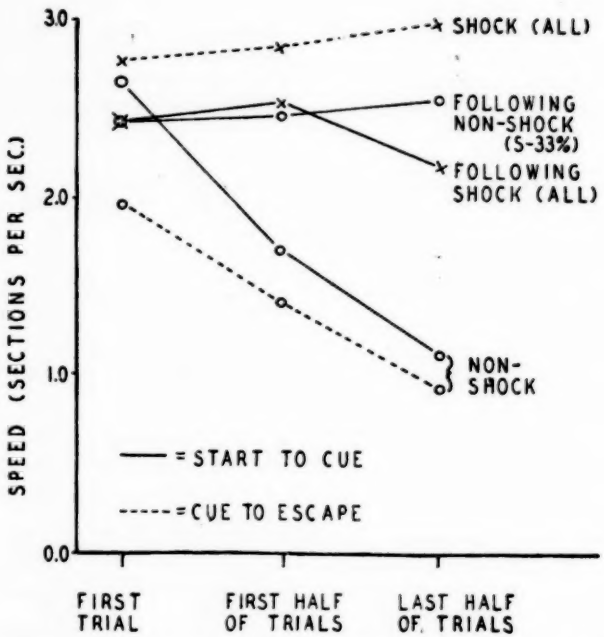


FIG. 2. Mean speeds on the first test trial and on the first and last halves of test trials. The curves show the speed of running to the cue on trials following shocks for all shocked animals, on trials following the non-shock trials in the shock-33% condition, and on all trials in the non-shock condition. They also show the speed of running from cue to escape on all shock trials and on all trials in the non-shock condition.

means for the various conditions of shock are all in the direction of longer running for shocked animals and are highly reliable for strong shock and shock every third trial.

b. *Effects of strong shock vs. weak shock.*—Referring again to Tables I and II, it may be seen that running to the cue was faster and continued for more trials when it was punished by strong shock than when punished by weak shock. For both measures, the means for strong shock were reliably greater than those for weak shock.

These facts indicate that punishment facilitated performance of the punished act and that this facilitation increased with the intensity of the punishing stimulus.

Evidence of inhibition by punishment

Despite the above facts, there was evidence in the behavior of the shocked animals which indicated that the shock not only facilitated the running tendency but also produced a conflicting tendency *not* to run into the shock area.

a. *Effects of shock in shock-33% condition.*—Running to the cue, in animals shocked at the cue on every third trial only, slowed down immediately following shock trials and speeded up during the intervening non-shock trials. The mean speed for all trials following shock is very reliably less (significant beyond the one percent level) than for trials immediately preceding shock.[6] This decrease in speed following the occasional shock trials, by comparison with speed following the non-shock, shows that the shock had inhibitory effects upon the act of running to the cue. It has already been shown (cf. Tables I and II) that, by comparison with no shock at all, the occasional shock facilitated that act. It may be said, therefore, that the shock had both facilitatory and inhibitory effects at least in the shock-33% condition.

b. *Effects of repetition of trials.*—Despite the facilitatory effects of the punishing shock, all the animals in all the shock groups as well as the non-shock eventually slowed down and stopped running to the cue long enough to meet the 10-sec. stopping criterion. Referring again to Fig. 2, it may be seen that the speed of running to the cue on trials following the punishing shock was less for the second half of trials run than it was for the first half. The mean difference in speed for the two halves is statistically significant beyond the three percent level. That this eventual slowing down after repeated shocks was an inhibitory consequence of the shock is suggested by the fact that the speed of running following the *non-shock* trials of animals shocked only on every third trial did *not* decrease with repetition of trials (cf. Fig. 2).

It is important to note, in consideration of 'interference' interpretations of the inhibitory effects of punishment, that for shocked animals the slowing down from start to cue accompanied a progressive increase in speed of running forward from cue to escape in response to the shock (cf. Fig. 2). The mean speed from cue to escape was significantly greater on the second half than on the first half of the shock trials (significant beyond the five percent level).

c. *Location of stopping and conflict behavior.*—The eventual decrease in the measured speed of running to the cue was due more to an increase in the occurrence and duration of stopping than to slower running. In shocked

[6] The analyses of the mean differences here and in the following paragraphs are based on distributions of the differences for individuals.

animals this tendency to stop was stronger, i.e., was more frequent, *just in front of the cue* than anywhere else in the alley. The greater frequency of stopping in front of the cue (sections 4–5) than near the start (sections 1–2) is significant beyong the one percent level. Stopping in the area from cue to escape was extremely rare in shocked animals even when shock was omitted in the shock-33% condition; for all individuals in the shock groups it was less frequent than stopping in front of the cue. All non-shock animals, on the other hand, stopped most frequently near the escape cage.

It was observed that the stopping of shocked animals in front of the cue was almost always accompanied by either 'freezing' or slight but obvious vacillatory movements followed by dashing forward to the escape cage. Stopping and then spurting past the cue eventually became so pronounced that visitors in the laboratory were inclined to believe that the experimenter shocked the animals by mistake on non-shock trials in the shock-33% condition. Such behavior did not occur in the non-shocked animals.

It should be noted that, in terms of sign-gestalt theories, stopping in front of the cue but running from cue to escape even when shock was omitted shows that the animals 'knew' the shock was located at the cue and were avoiding it; yet running into the shock area tended to continue for many trials after they acquired this knowledge.

These facts indicate that the punishment both facilitated and inhibited the punished act, i.e., it produced a conflict between tendencies to run to the escape cage and not to run to the cue.

The influence of frequency of punishment

In Tables I and II it may be seen that the running tendency of animals shocked on every trial did not differ significantly, nor in any consistent direction for the two measures, from that of animals shocked on every third trial only. It is possible that the different frequencies of shock did not have an differential influence. It has already been shown, however, that the inhibitory effects of the shock eventually increased with its repetition despite the facilitatory effects. This suggests that both inhibition and facilitation may have increased with the frequency of shock and that the counterbalancing of these opposing effects minimized measurable differences between shock every trial and shock every third trial.

Interpertation

On the basis of theories consistent with the general principle of inhibition by punishment it was expected that the punishment at the cue would inhibit rather than facilitate running to the cue. The fact is that the most powerful effect of the punishment was, for many trials at least, facilitation rather than inhibition of that act, and the facilitation increased with the intensity of the punishment. This suggests that punishment does not necessarily inhibit punished acts and that these theories need to be extended or stated more precisely to account for the experimental facts. It is possible to 'explain' the results in terms of sign-gestalt-expectancies by inferring from the observed behavior which of several possible expectancies must have operated to produce responses appropriate to them; but it is difficult, if not impossible, to derive these expectancies from expectancy theories as presently formulated.

On the other hand, the fact of facilitation by punishment was predicted in terms of S-R contiguity and drive-reduction theories including the fear hypothesis. This prediction, however, involved the assumption that the punishment would have no inhibitory effects. The fact that inhibition was observed suggests that a modification of these theories is required before they can provide an adequate account of the present results.

It is proposed that interpretation of the apparently 'inconsistent' effects of punishment in this and other experiments (Postman, 5) can be furthered by elaborating Mowrer's (4) most recent version of the fear hypothesis as follows:

1. Fear is a *central* rather than peripheral reaction [7] to noxious stimuli such as shock and is learned in accordance with Pavlovian (9) principles of conditioning by contiguity; whereas *sequences of locomotor responses* such as running are learned in accordance with principles of reinforcement by drive-reduction.

2. As a central process, fear directly mediates those S-R events usually regarded as indicative of fear, e.g., startle responses to sudden intense stimulation. Therefore, *a sudden increase in fear tends to elicit a startle response*, the strength of the startle being a positive function of the rate of increase. This response is an innate, short-latency 'reflex' of very short duration and is not subject to 'extinction' (Hunt and Landis, 10). While fear *per se* may evoke any of the possible locomotor responses in accordance with learning principles, an *increase* in fear always tends to evoke a startle.

[7] See Morgan (8) for various physiological theories of emotion, especially the Cannon-Bard thalamic theory.

It may be deduced from general S-R theory and the fear hypothesis, as modified above, that the various effects of punishment on punished acts are mediated by conditioned fear. The nature and amount of the effects depend upon the strength of the fear and upon the responses it tends to evoke. The effects of punishment are not mediated by the learning of 'anticipatory' locomotor responses to the punishing stimulus, because, unlike fear, such responses cannot be conditioned (and can generalize little, if at all) to the cues for the punished act.

An analysis in these terms of the present experimental results is suggested in the following paragraphs.

Facilitation by punishment

On the initial test trials, the fear conditioned to the alley during preliminary training motivated the animals to run and escape from the alley. On all test trials under all the test conditions this running response to fear continued to be rewarded by fear-reduction upon escape into the escape cage. But the punishing shock strengthened the conditioned fear throughout the alley, while fear was extinguished in the absence of shock. These changes in the strength of the fear-drive motivating the running response produced the observed speeding up of running for several trials in shocked animals in contrast with the slowing down in non-shocked.

Running was facilitated more by strong shock than by weak shock because the strength of the fear-drive increased with the intensity of the shock.

Inhibition by punishment

After many repetitions of shock at the cue, the strength of fear became greater just in front of the cue than

near the start, i.e., fear increased upon approach to the cue. This increase in fear eventually became rapid enough to evoke an effective startle and momentarily stop the running. The *rate* of increase in fear, and hence the observed inhibition of running, increased with repetition of shocks and was greatest just in front of the cue.[8] It did not increase with the intensity of the shock,[9] so the strong shock was no more inhibitory than the weak shock. The reason there were no significant differences between the observed behavior of animals shocked every trial and those shocked on a third of the trials only is that *both* the fear-drive strength and the effectiveness of the startle response increased with the frequency of shock; these opposing influences counterbalanced each other and minimized measurable differences between the two conditions.

There was no inhibition by punishment until after many trials. This was because, during the 18 preliminary training trials, the fear conditioned by shock was of approximately equal strength throughout the alley; but a tendency to *stop* being afraid in the escape cage, i.e., a non-fear response, was conditioned more strongly near the escape than near the start. Therefore, on the first test trials, as the animals ran from start to escape (via the cue) their fear decreased. In support of this proposition is the fact that, on the first test trials, non-shocked animals ran more slowly near the escape than near the start of the alley (cf. Fig. 2). Because of the counteracting effect of this initial decrease in fear as the animals ran toward the escape, many shocks at the cue were needed before an increase in fear accompanied approach to the cue. There could be no inhibition until such an increase occurred and became rapid enough to evoke an effective startle response.

Whenever a startle stopped the running, a tendency to remain immobile, i.e., to 'freeze,' was reinforced by fear-reduction to the pattern of visual, proprioceptive and other stimuli resulting from not running. The fear-reduction was due to the cessation of many of the fear-producing stimuli, i.e., the stimuli involved in running to the cue. The conflict behavior in front of the cue was due to the development of the tendency to freeze in opposition to the initially dominant tendency to run in response to fear. Through the process of discrimination, freezing eventually outcompeted running in front of the cue so that all animals reached the 10-sec. 'stopping' criterion.

[8] Statements concerning differences in the strength of conditioned fear in the various parts of the alley are derived from Hull's (11) quantitative assumptions concerning the strength of *S-R* connections.

[9] A study by Brown (12) of spatial gradients of avoidance, here assumed to be based on the strength of conditioned fear, showed that the height (strength of fear) but not the steepness (rate of increase in strength of fear) of the gradients increased with the intensity of the shock being avoided.

Summary

It is implied by 'interference' theories in general and the fear hypothesis in particular that acts motivated by fear will not be inhibited but rather will be facilitated by punishment, when they are compatible with the responses to the punishing stimulus. Furthermore, the facilitation will increase with the intensity and frequency of occurrence of the punishing stimulus. The present experiment established conditions designed to test this proposition as against the view that punishment necessarily tends to inhibit punished acts. The results showed that:

1. The punishment facilitated the punished act, and the facilitation increased with the intensity of the punishing stimulus. There was no evidence of inhibition until after many repetitions of the punishment.

2. Despite its facilitatory effects, the punishment eventually inhibited the punished act, and the inhibition increased with continued repetitions of the punishment.

3. The observed effects of punishment did not differ significantly for subjects punished on every trial as compared with those punished on every third trial only.

An interpretation of the results in terms of a modification of the fear hypothesis was suggested.

(Manuscript received for immediate publication December 20, 1948)

References

1. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
2. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1946.
3. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
4. Mowrer, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harvard Educ. Rev.*, 1947, **17**, 102–148.
5. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
6. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
7. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
8. Morgan, C. T. *Physiological psychology.* New York: McGraw-Hill, 1943.
9. Pavlov, I. P. *Conditioned reflexes.* (Trans. by C. V. Anrep.) London: Oxford Univ. Press, 1927.
10. Hunt, W. A., & Landis, C. The overt behavior pattern in startle. *J. exp. Psychol.*, 1936, **19**, 309–315.
11. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
12. Brown, J. S. Gradients of approach and avoidance responses and their relation to level of motivation. *J. comp. Psychol.*, 1948, **41**, 450–465.

ACCURACY OF VISUAL INTERPOLATION BETWEEN CIRCULAR SCALE MARKERS AS A FUNCTION OF THE SEPARATION BETWEEN MARKERS [1]

BY M. LEYZOREK [2]

The Johns Hopkins University

Introduction

The measurement of distance with any physical scale (e.g., a yardstick) involves two psychological problems. One is that of accurately counting scale markers up to the one nearest the point being located. The other is estimating the position of the point relative to the scale marks between which it falls. The experiments reported here are concerned with the latter process.

Status of earlier work on this problem.—The standard psychological literature is rich in studies of a certain kind of localization, viz., the bisection of lines and spaces between lines, but there has been almost no work until recently on more general aspects of this problem. The small number of references in the appended bibliography probably does not adequately do credit to the large volume of research now being done on the psychology of scales. Much of this work, particularly that dealing with scales for instrument dials, is in progress under the auspices of the Army Air Forces research program in both military and university laboratories (e.g., Aero-Medical Laboratory, Wright Field; Princeton University, New Jersey). Unfortunately, much of this research has not yet been reported in the standard professional journals.

Origin of this study.—This report is one of a series of studies completed (**1, 2, 4**) or in progress at this laboratory on the characteristics of scales. These experiments fall into two principal groups. One group of experiments is concerned with the psychological aspects of counting scale marks. It deals with the relationship between speed and accuracy of counting and (a) the number of scale divisions, (b) the patterning of scale markers, (c) the frequency of numbered markers vs. unnumbered markers, and (d) the numeral assigned to scale markers.

The other set of studies, of which the experiment reported here is one, deals with the problem of visual interpolation between scale marks. This kind of spatial localization problem is characteristic of the use of all types of scales, whether linear (e.g., meter stick) or circular (e.g., as in protractors, instrument dials). This kind of estimation is also made along any radius of a polar coordinate plot supplied with circular scale marks, as illustrated in Fig. 1.

Recently, Reed and Bartlett (**6**) studied the accuracy of localization of a point as a function of its position between two scale markers on a polar

[1] This experiment was carried out under Contract N5ori-166, Task Order I, between Special Devices Center, Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-52, Project Designation No. NR-784-001 under that contract.

[2] The guidance and suggestions of Drs. A. Chapanis, J. W. Gebhard, and W. R. Garner concerning the design of this experiment and the processing of its data are deeply appreciated. Many thanks are due the engineering and technical staff of the Field Laboratory of the Systems Research Project for their assistance with equipment and materials, and the Navy personnel, who participated as subjects, for their fine cooperation.

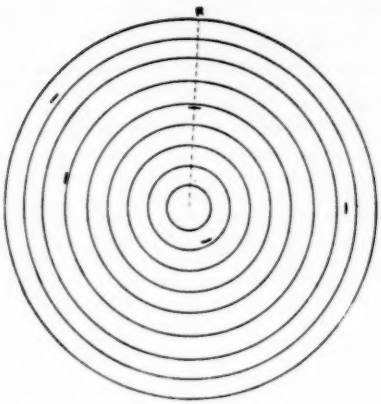


FIG. 1. Schematic representation of a 10-in. polar coordinate plot with ½-in. scale intervals. The short arcs are the points whose positions must be estimated by visual interpolation between scale markers. Interpolation must be made from the center outward along an imaginary radius such as R.

coordinate plot (See Fig. 1). They found that accuracy of localization varies complexly with the relative position of the point and the scale markers. In general, points near scale markers and those near the center of the interval (midway between rings) were located more accurately than others. In another study Bartlett, Reed, and Duvoisin (1) investigated the accuracy of distance estimation in a polar coordinate display as a function of three different numerical scales assigned to the same pattern of scale marks. This study was designed to elicit the contribution made by the numerals assigned the scale markers to the errors of estimation. It was found that the evidence for the effect of numerical scale on accuracy of interpolation was somewhat equivocal but strongly suggested some relationship. Earlier, Chapanis (2) had studied the same problem. He tested eleven different numerical scales on a polar coordinate plot with four fixed scale markers and found that the accuracy of distance estimation was significantly affected by the numeral assigned to the scale unit. Since all of these studies were done with scales of fixed size, they logically suggested an investigation of the factor of scale size.

Purpose of this experiment.—The purpose of this research, then, was to investigate the relationship between accuracy of visual interpolation in a polar coordinate display and the physical size of the scale intervals provided.

APPARATUS, SUBJECTS AND PROCEDURE

Apparatus

The visual problem.—The visual problem presented to the subject appeared as a fairly dark, unsaturated blue circular field covered with a sharply defined pattern of thin, black, equally-spaced, concentric rings. These served as circular scale markers (see Fig. 1). Interspersed randomly among these rings (and concentric with them) were short, diffuse white arcs. These were to be located by visual interpolation between the appropriate scale markers.

Component items.—The displays used in this experiment consisted of four items:

1. Blue-printed arrays of the short, white arcs used for localization purposes. These appeared in groups of 5 or 15 on blue circular fields either 10″ or 20″ in diameter. The arcs were about 6° in angular width (with respect to the total circumference of the plot) and were 1.1 mm. thick (measured along a radius of the circular field).

2. Circular scales drawn in India ink on sheets of translucent draftsman's linen. The lines were approximately 0.35 mm. thick. Fourteen different scales were used, of which half were 10″ overall in diameter and the other half 20″ overall in diameter. The seven scale sizes tested in the 10″ size were ⅛″, ¼″, ½″, 1″, 2″, 4″, and 5″ (e.g., on the ½″ scale, the concentric rings or scale markers were ½″ apart. See Fig. 1). In the 20″ size, the scales tested were 1″, 2″, 4″, 5″, 6″, 8″, and 10″.

3. Opaque masks of cardboard with 10″ and 20″ diameter openings, respectively.

4. A large, square sheet of transparent plastic.

TABLE I

DESIGN OF THE TWO EXPERIMENTS

The numbers below the scale intervals are the serial orders in which each scale was tested with each subject. The arabic numerals refer to trials in the first experiment with the 10″ display, the roman numerals to trials in the second experiment with the 20″ display.

Subject	Display	Scale Intervals Tested									
		¼″	½″	¾″	1″	2″	4″	5″	6″	8″	10″
Pas	10″	1	5	7	3	6	4	2			
	20″				IV	VI	I	V	II	III	VII
Thu	10″	3	6	4	7	5	2	1			
	20″				I	III	VI	IV	VII	II	V
Hag	10″	5	2	6	4	1	7	3			
	20″				V	II	IV	VII	I	VI	III
Has	10″	4	1	3	5	2	6	7			
	20″				II	IV	III	VI	V	VII	I
Gra	10″	2	7	1	6	4	3	5			
	20″				III	VII	II	I	VI	V	IV
Gou	10″	6	3	2	1	7	5	4			
	20″				VII	I	V	II	III	IV	VI
Bar	10″	7	4	5	2	3	1	6			
	20″				VI	V	VII	III	IV	I	II

Assembly of items.—These items were combined to form the complete display by first setting a blue-printed array of localization points on the transparent surface of a specially designed projection table, illuminated from below by a 200-watt lamp. Over the blue-printed sheet was placed one of the translucent scale overlays. This was then covered by a cardboard mask of the appropriate size and finally by the sheet of clear plastic. The plastic served to hold down the complete display, eliminate wrinkles, and to permit leaning on the display without disorienting it. The correct alignment of these parts was insured by fitting the component parts of the display over pins projecting from the surface of the projection table which matched corresponding holes in the blue-print sheets, the scale overlays, the masks, and the covering sheet of plastic.

Experimental Design

Two separate experiments were performed, the first with the 10″ display, the second with the 20″ materials. Each experiment was designed as a Latin Square (cf. Table I). As is evident from an inspection of the table, the scales were tested in a different order for each subject without repetition of any sequence of tests. Insofar as possible, the patterns of points were randomized in the same fashion, although there were a few unavoidable repetitions. These are considered to have had no effect on the results, since no target pattern was ever repeated for any subject with the same scale.

In this way, an attempt was made to balance several possible sources of bias in the experimental data: (a) learning, practice, or other serial effects due to the order in which the various scales were presented to the subjects, (b) memory effects due to familiarity with a given point pattern, and (c) effects due to insufficient sampling of the circular field for each scale tested (interpolation might be more difficult in some quadrants than in others).

Subjects

The subjects were seven experienced naval radar operators, familiar with the process of visual interpolation in circular scales on polar coordinate plots.

Procedure

A numerical value of 1000 yards was assigned to each scale interval regardless of size, and the subjects were asked to make their estimates of the position of each point in terms of this unit value. Earlier research (a) has shown that subjects can make precise visual interpolations when

this number is assigned to the scale interval. The actual position of each point was carefully measured and its nominal position was calculated in terms of each scale before any data were taken.

The subjects were instructed to take each point in turn, going in a clockwise direction, and estimate its position as accurately as possible. No time limit was imposed and the need for accuracy was emphasized. No restriction was placed on the subject's viewing position. (The viewing distance ranged from 12 to 24 in.) A total of 20 items (in some combination of 5 and/or 15-item groups) was presented to each subject during one trial. One trial was given per subject per scale tested. The average time consumed per trial was 35 min. with a range from 20 to 45 min.

Measurements

By contrast with the data of the classical method of average error, which are deviations from a fixed standard, the raw data of this experiment are deviations from a standard which varied numerically for each judgment made by the subjects. For each judgment made by the subjects, an error deviation was obtained by subtracting the true position of each point from its estimated position. Each error deviation was then expressed as a proportional error of estimation, i.e., as a percentage of the scale interval. By their arithmetic signs these proportional error scores identified the direction (i.e., whether over or underestimation), as well as the relative magnitude of the estimation errors.

These proportional error scores were treated statistically in three ways. (a) When they were averaged for each trial (20 estimates) without regard to sign, the average arithmetic error per trial was obtained. This datum is the average deviation around zero and it provides an indication of the magnitude of the estimation errors made by the subject, on the average, with each scale used.

TABLE II

DATA FROM THE EXPERIMENT WITH THE 10″ DISPLAY

The upper entry in each cell is the average arithmetic error; the middle entry is the average algebraic error (constant error); the lower entry is the error variability (σ) calculated around the constant error. N in each cell is 20. Errors are expressed as percentages of the scale intervals.

Subject	Scale Intervals							Mean
	¼″	½″	¾″	1″	2″	4″	5″	
Pas	12.825 +3.325 14.308	8.400 +3.600 8.760	3.700 +1.050 4.350	3.789 +1.737 4.689	3.225 +1.275 3.816	2.100 +1.580 2.214	4.145 +1.995 4.742	5.455 +2.080 6.126
Thu	10.525 −7.025 10.408	9.475 +4.725 10.757	5.400 +4.750 5.190	5.375 +4.275 4.608	5.610 +4.390 4.995	7.435 +7.435 4.517	4.785 +3.895 3.930	6.944 +3.206 6.344
Hag	25.275 −4.975 30.517	12.525 +4.025 17.733	5.075 +3.275 5.647	5.200 +4.400 4.535	7.030 +5.100 6.849	8.535 +6.155 7.230	6.260 +1.610 7.065	9.986 +2.799 11.368
Has	17.100 +6.300 20.478	10.550 −1.250 14.039	3.725 +1.775 4.755	4.800 +1.150 5.723	4.655 +0.755 5.860	5.845 +4.525 4.704	3.980 −0.950 4.858	7.236 +1.758 8.631
Gra	11.800 −2.450 13.151	4.250 0.000 5.572	5.125 −2.325 6.221	2.175 −0.325 2.929	2.890 +1.320 3.151	2.265 +1.625 2.621	1.705 +0.065 2.369	4.316 −0.299 5.145
Gou	31.500 +6.750 36.228	7.450 +1.050 9.769	2.950 +1.350 3.719	3.925 +2.925 4.051	3.455 +2.711 2.974	3.915 +3.915 2.589	2.820 +0.800 5.351	8.002 +2.786 9.240
Bar	17.600 +5.400 19.984	7.600 +1.900 8.959	3.200 +0.600 4.042	2.925 −0.575 4.305	3.110 +0.880 3.660	6.795 +4.345 6.658	3.805 +1.245 4.540	6.434 +1.971 7.450
Mean	18.089 +1.046 20.725	8.607 +2.007 10.798	4.168 +1.496 4.846	4.027 +1.941 4.405	4.282 +2.347 4.472	5.270 +4.226 4.362	3.929 +1.237 4.694	

TABLE III

DATA FROM THE EXPERIMENT WITH THE 20″ DISPLAY

The upper entry in each cell is the average arithmetic error; the middle entry is the average algebraic error (constant error); the lower entry is the error variability (σ) calculated around the constant error. N in each cell is 20. Errors are expressed as percentages of the scale intervals.

Subject	Scale Intervals							Mean
	1″	2″	4″	5″	6″	8″	10″	
Pas	3.475 −0.275 4.414	4.100 +2.700 4.323	2.550 +1.200 2.981	2.335 +1.215 2.621	3.162 +1.762 3.458	4.870 +4.220 3.559	3.445 +0.785 3.804	3.419 +1.658 3.594
Thu	7.857 +7.285 5.317	6.000 +5.100 5.787	3.575 +1.025 4.039	3.390 +0.610 4.071	4.000 +3.370 4.592	5.420 +2.610 6.236	6.152 +5.647 3.998	5.199 +3.664 4.863
Hag	5.075 −0.925 6.809	4.157 +3.315 4.030	7.261 +4.690 6.620	8.062 +5.300 8.414	6.447 +5.019 5.657	7.340 +6.620 5.235	3.990 +1.700 4.639	6.047 +3.674 5.915
Has	4.473 −2.158 5.932	3.947 +0.895 5.129	4.595 +2.309 4.934	3.590 +2.543 4.372	5.995 +1.925 6.365	11.300 +6.233 15.247	2.905 +1.085 3.399	5.257 +1.833 6.483
Gra	6.225 +1.175 7.782	2.825 −1.925 3.348	2.165 +0.715 2.583	2.467 +1.390 2.791	2.280 −0.700 2.897	2.320 +1.070 2.559	1.585 −0.465 2.180	2.838 +0.180 3.449
Gou	3.025 −0.675 4.136	3.275 +1.425 3.995	2.452 +1.881 2.193	3.270 +1.240 4.100	3.200 +2.440 2.861	4.189 +3.378 3.039	4.255 +1.085 5.085	3.381 +1.539 3.630
Bar	5.300 +3.900 5.200	1.850 −0.900 2.375	3.715 +1.615 4.192	2.757 +1.729 2.728	3.600 +1.714 3.887	5.171 +4.724 3.276	3.585 +0.995 3.911	3.711 +1.968 3.653
Mean	5.061 +1.190 5.656	3.736 +1.516 4.141	3.759 +1.919 3.935	3.696 +2.004 4.157	4.098 +2.219 4.245	5.801 +4.122 5.593	3.702 +1.547 3.859	

(b) The constant error per trial was furnished by the algebraic mean of each 20 error scores. From this datum it can be determined whether the subjects consistently over or underestimated the position of each point when using a given scale. (c) Each subject's variability was determined by calculating a standard deviation for the error scores of each trial about his own mean (constant error). These three categories of data appear in Tables II and III as the upper, middle, and lower entries, respectively, of each cell in the tables.

The average arithmetic error data have been selected as the primary criteria of estimation accuracy, but it should be clear that the average arithmetic error is a function both of the constant error and error variability (standard deviation) around the constant error. Since the constant errors are shown later not to differ significantly from each other, the major conclusions of these experiments are unchanged when error variabilities are analyzed instead. From the standpoint of the practical application of these data, the average arithmetic error is most easily interpreted: it shows 'how far off' the estimates were, on the average, with each scale.

Since precautions were taken to eliminate errors due solely to miscounting scale marks up to the one beyond which interpolation began, there were only a few such errors in the final mass of data. These were discarded in the computation of the average error values.

RESULTS

Average errors.—The average arithmetic error of estimation (expressed in percent of the scale interval) is presented in Fig. 2 as a function of the

physical size of the interval. From this function it is clear that the proportional errors of interpolation are relatively large for the smallest scale interval ($\frac{1}{8}''$) but decrease rapidly as the size of the interval increases to $\frac{1}{2}''$. For scale intervals larger than $\frac{1}{2}''$ there appears to be no change in the size of this error.

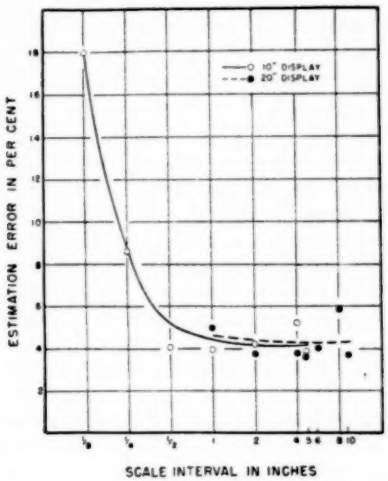


FIG. 2. Average arithmetic error of visual interpolation (expressed as percent of the interval) as a function of scale interval size

The significance of these error differences was tested by the analysis of variance technique with the results shown in Tables IV and V. Differences in performance on the 10" display due to factors in the scales themselves are significant at well below the one percent level, while the variance due to differences between subjects is barely significant at the five percent level. In the 20" display, the scales did not differ significantly, while the differences between the subjects were significant at the one percent level.

Constant errors.—The constant errors of estimation (average algebraic errors) listed in Tables II and III were also analyzed with the same technique. No significant differences were found between scales in terms of constant errors of interpolation either with the 10" or 20" data. Differences between subjects were again significant in both experiments at between the one and five percent levels.

It is interesting to note that the constant errors are—with few exceptions—positive in sign. Taken as a whole, this is a statistically significant finding. It reveals a small but consistent tendency for subjects to overestimate the positions of the points, i.e., to judge them as being more peripheral than they were. The data of these experiments do not, unfortunately, provide an explanation for this finding.

Subjects' variability.—As has already been stated, the average arithmetic error for a set of measurements is a function of both the constant error and the variable error around the constant error. Since the constant errors do not differ significantly from each other, analyses involving measures of error variability should yield the same results as were obtained with the average arithmetic errors. This is essentially what was found. The analysis of variance of the standard deviations for the 10" data (lower entries in each cell of Table II) shows the differences between scales to be highly significant (well below the one percent level of confidence) and the differences between subjects somewhat less significant (slightly below the one percent level). The analysis of variance of the 20" data shows only the differences between subjects to be significant (at the one percent level).

Individual differences in estimation ability.—The significance of the differences between subjects in respect to their errors (both constant and variable) can be interpreted as re-

TABLE IV

ANALYSIS OF VARIANCE OF AVERAGE ARITHMETIC ERRORS IN THE EXPERIMENT WITH THE 10″ DISPLAY

The basic data for this analysis are the uppermost entries in each cell of Table II.

Source of Variance	Degrees of Freedom	Sums of Squares	Estimated Mean Square	F
Between scales	6	1135.2033	189.200	19.56*
Between subjects	6	138.8465	23.141	2.39**
Between successive trials	6	49.9687	8.328	.86
Discrepancy	30	290.1156	9.671	
Total	48	1614.1341		

* Significant at the 1% level
** Barely significant at the 5% level

flecting real and possibly predictable individual differences between subjects in their ability to perform the visual interpolation required in these experiments. The fact that these experiments were done separately (i.e., data were collected first from the 10″ display, then from the 20″ display) makes it possible to test this assumption in a crude way.

Of the 10 scales tested, four (i.e., the 1″, 2″, 4″, and 5″ scales) were tested on both the 10″ and the 20″ displays. In terms of these four scales, the two displays, although different in overall size, present very nearly identical visual tasks. Thus, if the factor of overall display size is ignored, the two series of judgments made by each subject (one series in each of the two displays with each of the four scales mentioned above) constitute a test and retest of the same psychological function.

Three rank-difference correlations were computed between the subjects' performance in the 10″ display vs. that in the 20″ display. These three correlations were based on the subjects' average error, constant error, and variability when using the 1″, 2″, 4″, and 5″ scales. The subjects' average performance score with these four scales was used in establishing the rank orders. Thus, $N = 7$. The data basic to these correlations were extracted from Tables II and III. The correlations appear in Table VI. Even taking into account the crudeness of the rank-difference statistic and the small number of cases contributing to this analysis, these corre-

TABLE V

ANALYSIS OF VARIANCE OF AVERAGE ARITHMETIC ERRORS IN THE EXPERIMENT WITH THE 20″ DISPLAY

The basic data for this analysis are the uppermost entries in each cell of Table III.

Source of Variance	Degrees of Freedom	Sums of Squares	Estimated Mean Square	F
Between scales	6	29.3922	4.899	2.10
Between subjects	6	62.1300	10.355	4.45*
Between successive trials	6	11.2992	1.883	.81
Discrepancy	30	69.8705	2.329	
Total	48	172.6919		

* Significant at the 1% level

TABLE VI

RANK-DIFFERENCE CORRELATIONS BASED ON THE PERFORMANCE OF 7 Ss ON IDENTICAL SCALES IN TWO EXPERIMENTS

Performance Factor	Rank-Difference Correlation	Standard Error
Average error	0.750	0.1876
Constant error	0.857	0.1138
Variability	0.714	0.2100

lations are high enough, when supported by the previously demonstrated significant differences between subjects, to suggest the following conclusions:

1. Differences between subjects in their ability to perform the kind of visual estimation studied here are reliable.

2. Relative levels of performance can be predicted from one test to another.

3. If, at the outset, the differences between subjects are assumed to be reliable and the similarity between the 10″ and 20″ displays is in question, then this test becomes a measure of such similarity. From the correlations obtained, it can reasonably be assumed that the 10″ and 20″ displays presented very nearly the same visual task.

Similarity of 10″ and 20″ display data.—From Fig. 2 it is apparent that the $\frac{1}{8}$″ and $\frac{1}{4}$″ scales were primarily responsible for the significance of the differences between scales found in the 10″ display data. On this basis, no further test was made of differences between the $\frac{1}{8}$″ and $\frac{1}{4}$″ scales versus the others tested with the 10″ display. Following the same line of reasoning, since the data obtained with the 20″ display did not show significant internal differences (with respect to the scales) and were of the same order of magnitude as those obtained with the 10″ display (other than the $\frac{1}{8}$″ and $\frac{1}{4}$″ data, see Fig. 2), no attempt was made to analyze for the significance of differences between the two sets of data.

Absolute errors.—The interpolation error data of Fig. 2 are expressed as proportions of the intervals in which the estimations were made. In Fig. 3 these errors are stated as absolute errors in millimeters as a function of the scale interval expressed in the same units. Note from Fig. 3 that these errors of interpolation are virtually constant as the scale interval increases in size from $\frac{1}{8}$″ to $\frac{1}{2}$″. Then they increase linearly as the size of the interval increases beyond $\frac{1}{2}$″. This fact, plus the information on proportional error previously cited, may be of considerable value for the design of scales in the use of which visual interpolation plays an important part.

It is of some interest that the average absolute errors made with the $\frac{1}{8}$″, $\frac{1}{4}$″, and $\frac{1}{2}$″ scales correspond to spatial discrimination of the order of 4 minutes of arc (based on an average viewing distance of 18″). While this is not an impressive figure when compared to the discriminations achieved under ideal conditions in psychophysical experiments, these are very precise judgments in terms of the kind of visual problem presented here.

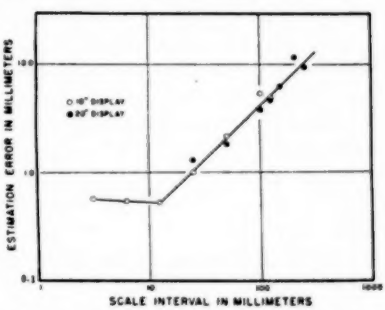


FIG. 3. Average arithmetic error of visual interpolation (expressed as absolute error in mm.)

Cyclical error variation.—In several of the studies previously cited (1, 2, 4, 6), it was found that the interpolation errors varied complexly. In one study Chapanis (2) found that the position of the point with respect to the scale markers was consistently overestimated in the first half of the interval. In the second half its position was consistently underestimated, thus giving rise to a cyclical error function which was at a minimum at about the center of the interval. Bartlett, Reed, and Duvoisin (1) also found some evidence of cyclical variation in the constant errors of estimation, but without the regularity or consistency of the cycles in the Chapanis study.

To explore this phenomenon further, the data of this experiment were plotted to show the average error, constant error, and variability of estimation as functions of the position of the point between the two scale markers. In general, these plots, while they resemble slightly those of the Chapanis study (2), are not clear-cut support for any hypothesis about cyclical variation of constant error in interpolation between scale markers.

Discussion

It is obvious that generalization from the findings already stated must be limited to the extent to which certain parameters affect them. Some of these are: the ratio of point thickness to interval size; the ratio of the thickness of the point located to the thickness of the scale marks; and the ratio of marker thickness to scale interval size. Any or all of these ratios may affect the accuracy of visual interpolation. This, then, can be considered a preliminary study so far as complete generalization on the subject is concerned. For a number of practical problems, however, it provides information which, when integrated with the results of other studies cited, may help in the design of better scales for a variety of uses.

There is some indication from other studies (3, 5) that the principal finding of this experiment may apply to circular scales (as on instrument dials or protractors) as well as to the polar coordinate type used here. This study has shown that $\frac{1}{2}''$ is a critical value for the size of scale units in which visual interpolation is attempted. In general, the proportional errors of estimation have been found to be large for scale units smaller than $\frac{1}{2}''$ and relatively constant for larger scale units, while absolute errors have been found to be approximately constant for scale units smaller than $\frac{1}{2}''$ but to increase linearly for units larger than $\frac{1}{2}''$. By suitable extrapolation from the data of the studies cited above (3, 5) it appears that the same general relationship holds for interpolation in circular scales.

Summary

1. The accuracy of visual interpolation between the circular scale markers of a polar coordinate plot has been studied as a function of the size of the scale interval. The interval sizes studied were $\frac{1}{8}''$, $\frac{1}{2}''$, $1''$, $2''$, $4''$, $5''$, $6''$, $8''$, and $10''$. The first seven of these scales were studied on a $10''$ diameter display while the latter seven (including duplications of the $1''$, $2''$, $4''$, and $5''$ scales) were tested on a display $20''$ in diameter.

2. The principal finding of this study is that the errors of interpolation (expressed as proportions of the intervals in which the estimations are made) are large for scale intervals $\frac{1}{8}''$ in size (18 percent), and decrease rapidly as the scale interval increases in size up to $\frac{1}{2}''$ (4 per cent). In-

creases in the size of the scale interval beyond $\frac{1}{2}''$ do not further reduce errors of estimation.

3. Expressed as absolute errors of estimation in millimeters, the errors of localization are constant for the three smallest scale intervals tested ($\frac{1}{8}''$, $\frac{1}{4}''$, $\frac{1}{2}''$) and increase linearly thereafter for larger scale units. For the larger scale units the Weber fraction is a constant.

4. Individuals differ significantly in their ability to perform this kind of visual interpolation.

(Manuscript received for immediate publication December 20, 1948)

References

1. Bartlett, N. R., Reed, J. D., & Duvoisin, G. Estimations of distance on polar coordinate plots as a function of the scale used. Systems Research, The Johns Hopkins University. Report No. 166-I-44. (In press)
2. Chapanis, A. Accuracy of interpolation between scale markers as a function of scale interval number. *Amer. Psychologist* 1947, 2, 346.
3. Grether, W. F., & Williams, A. C., Jr. Speed and accuracy of dial readings as a function of dial diameter and angular separation of scale divisions. Ch. 7 in: Fitts, P. M. (Ed.) *Psychological research on equipment design.* Washington: Gov't Printing Office, 1947, pp. 101–109.
4. Harriman, M. W., & Bartlett, N. R. Estimation of the relative position of a mark along a short line. (In press)
5. Kappauf, W. E., Jr., & Smith, W. M. A preliminary experiment on the effect of dial graduation and dial size on the speed and accuracy of dial readings. *Ann. N. Y. Acad. Sci.*, 1948. (In press)
6. Reed, J. D., & Bartlett, N. R. Unpublished data, Systems Research, The Johns Hopkins Univ.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	1	–	3	4	5	6	$1.25	$6.25
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	By subscription, $7.00						$1.25	——
		List price, Volumes 1 through 38							$247.50
		30% discount							74.25
		Net price, Volumes 1 through 38							$173.25

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year. Eighteen numbers affecting eight volumes are out of print in the entire series.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W. Washington 5, D. C.

Special Pre-Publication Offer to Readers of the Journal

The Encyclopedia of

CRIMINOLOGY

Edited by **Dr. Vernon C. Branham** ***and*** **Dr. Samuel B. Kutash**

THIS comprehensive work will serve the needs of all those concerned with the problems of crime and criminality. It is a compendium of existing facts and knowledge in criminology and deals at the same time with important current trends and departures in this highly complex field. This has been accomplished through the combined efforts of sixty-one outstanding specialists in a variety of disciplines, including:

PSYCHOLOGY	**PSYCHIATRY**	**SOCIOLOGY**	**MEDICINE**
LAW	**POLICE SCIENCE**	**EDUCATION**	**HISTORY**
PENOLOGY	**PHILOSOPHY**	**RELIGION**	**RESEARCH**

CONTRIBUTORS INCLUDE: ***Hon. Justice*** **Francis Bergan; Nathaniel Cantor,** ***Ph.D.;*** **Hervey Cleckley,** ***M.D.;*** **Marshall B. Clinard,** ***Ph.D.;*** **Victoria Cranford;** ***Senator*** **Thomas C. Desmond; Arthur N. Foxe,** ***M.D.;*** **G. I. Giardini,** ***Ed.D.;*** **Eleanor Glueck,** ***Ph.D.;*** **Sheldon Glueck,** ***Ph.D.;*** **Irving W. Halpern; Leland E. Hinsie,** ***M.D.;*** ***Hon.*** **Alexander Holtzoff;** ***Hon.*** **J. Edgar Hoover;** ***Warden*** **Lewis E. Lawes; V. A. Leonard,** ***Ph.D.;*** **Robert M. Lindner,** ***Ph.D.;*** **Edwin J. Lukas,** ***LL.D.;*** **Elio D. Monachesi,** ***Ph.D.;*** **Winfred Overholser,** ***M.D.;*** **Michael J. Pescor,** ***M.D.;*** **Walter C. Reckless,** ***Ph.D.;*** **Orlando F. Scott,** ***M.D.;*** **Robert V. Seliger,** ***M.D.;*** **Lowell S. Selling,** ***M.D.;*** **Herman K. Spector; Edwin H. Sutherland,** ***Ph.D.;*** **Negley K. Teeters,** ***Ph.D.;*** **Louis P. Thorpe,** ***Ph.D.;*** **Joseph G. Wilson,** ***M.D.;*** **Fritz Wittels,** ***M.D.;*** **Martin L. Wolf,** ***LL.D.;*** **Arthur L. Wood,** ***Ph.D.***

Publication: **June, 1949. Approx. 500,000 words.** **$12.00**

Publication date is June, 1949. If you send in your order before publication, you can get your copy for

ONLY $10.00

SPECIAL ORDER COUPON

PHILOSOPHICAL LIBRARY, ***Publishers***
15 East 40th Street, Dept. 188, New York 16, N.Y.

Enclosed is $_____ for _____ copy(ies) of THE ENCYCLOPEDIA OF CRIMINOLOGY, at the special Pre-Publication price of $10.00 per copy. The book(s) to be mailed to me immediately upon publication.

NAME _____

ADDRESS _____

(Expedite shipment by enclosing remittance)